Optimising Power System Reserve for Contingencies while considering Response Times

Josh Schipper

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

June 2019

ABSTRACT

This thesis develops an optimisation formulation for the efficient procurement of reserve. The formulation explicitly accounts for delay and ramp rates in reserve response, and the effect they have on keeping the grid frequency above minimum limits. This research is conducted as a part of GREEN Grid project, which is seeking to understand the impacts of Variable Renewable Energy (VRE), such as wind and solar, and other distributed resources on the New Zealand power system. After considering the impacts of wind generation on reserve requirements, a method is developed to compare reserve resources for contingencies more accurately, which can be applied to the scheduling and dispatch processes.

In a power system, contingencies like the loss of a generator or transmission circuit can create an imbalance, where generation is insufficient to supply load. Reserve is required to increase power output to stop the declining grid frequency, otherwise the system collapses. The amount of time available for reserve to respond is dependent on three factors: the size of the largest credible contingency, the inertia of the power system, and allowable frequency range. To increase the available time to respond, the size of the contingency is to be small, inertia should be maximised, and the allowable frequency range is to be wider. However, with increased penetrations of VRE, it is expected that the size of the largest credible contingency can increase and the inertia can decrease, thereby requiring greater response speed from reserve.

In electricity markets across the world, for the most part only capacity and price of reserve feature into decisions of optimal dispatch. Any transient features of reserve responses are considered too difficult for MILP methods. The literature, recognising the importance of response speed, has provided means of incorporating transient features, but have suffered from a lack of generality or insufficient computational performance. Therefore a new approach is developed that allows a wider range of responses to be optimised while being solved in a practical amount of time. It comes at the cost of greater complexity, and deviates from MILP. However it retains convexity like LP, which is beneficial for practical solve times. Further development is required for the inclusion of mixed integer variables and towards achieving its full potential of being implemented in real-time electricity markets.

The new formulation introduces the swing equation, which defines frequency dynam-

ics, into the optimisation formation. Reserve is limited to step and ramp responses that can be delayed. Frequency limits are applied at critical times and quadratic constraints are formed on account of ramped reserve that typifies responses from conventional power stations. This problem cannot be purely classified as a Quadratically Constrained Programming (QCP) problem, as the constraints change between being quadratic and linear depending on the location in the feasible solution space. Therefore this solution method is more accurately called Piecewise Quadratically Constrained Programming.

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:
Chapter 3 is adapted from a currently unpublished report for the GREEN Grid project, which is awaiting external review: Schipper, J., Wood, A., Edwards, C., and Miller, A. (2019), 'Recommendation for Ancillary Service Markets under High Penetrations of Wind Generation in New Zealand'

Please detail the nature and extent (%) of contribution by the candidate:
The PhD candidate did the research and wrote all sections of the thesis. The co-authoring supervisory team suggested research directions and offered criticisms which the candidate worked into the manuscript.

Certification by co-authors

If there is more than one co-author then a single co-author can sign on behalf of all. The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work, he or she wrote the text.

Name: *Associate Professor Alan Wood* Signature: *Alan Wood* Date: *28 June 2019*

CONTENTS

ACKNOWLEDGEMENTS

As the end draws near for my time at the University of Canterbury, I would like to thank those who have supported me and who have made it a fun experience. Firstly to my supervisors, Alan Wood, Allan Miller, Conrad Edwards, and Neville Watson, thank you for time, effort, and the knowledge gained.

Thank you to MBIE and Transpower for funding me through the GREEN Grid Project.

Thank you to those at the EPECentre for your support, discussions, battling through two office changes, for reviewing my work, and about life in Kazakhstan: Shreejan Pandey, Isabelle Le Quellec, Bill Heffernan, Richard Strahan, Sharee McNab, Ryan Van Herel, Nurzhan Nursultanov, David Santos-Martin, Timothy Crownshaw, Michael Hwang, Sue Bradley, Danica Nel, Ruth Payne, Kirsten Marsh, Linsey Mackenzie, Susannah Hawtin, Nicola Frederiksen, Marie-Claire Brehaut, Ian Mason, Andrew Lapthorn, Pat Bodger, and Radnya Mukhedkar.

To my fellow postgrads and technicians, thank you for your help, advice, and to solving optimisation problems: Luke Schwartfeger, Michael Campbell, Thomas Smart, Dougal McQueen, Euan McGill, Parash Acharya, Scott Lemon, Yanosh Irani, Lance Frater, Zhiyang Jin, Rahman Peimankar, Hantt Cao, Patrick Chen, Rabia Nazir, Diwakar Brunel, Alejandro Escamilla, Hae Guey, Michael Frampton, Blair Bonnet, Azy Hashemi, Andrew Berry, Ryan Cheyne, Rob Hockey, Muhammad Ramzan, Miguel Reis, Philipp Sueltrop, Ken Smart, Edsel Villa, Paul Agger, David Healy, Florin Predan, Mike Shurety, Jac Woudberg, Catherine Adamson, and Deborah Erueti. Apologies to anyone I have missed.

To those in industry, thank you for providing information that grounded my research: Richard Sherry, Charles Crystal, and Rowan Sinton.

To my past and current flatmates, friends, Christian Union and Cornerstone Church family, thank you for the encouragement, spiritual growth, and fun times. Especially to Cameron Oliver for an understanding ear in tough times.

To my family, Dad, Mum, Raoul, and Hannah, thank you for your love.

To God, I hoped my studies would provide time to understand the complexities of this World, I am disappointed I only touched the surface. *Soli Deo Gloria*

NOTATION

Acronyms

AGC Automatic Generation Control

AUFLS Automatic Under-Frequency Load Shedding

BESS Battery Energy Storage System

CCGT Combined Cycle Gas Turbine

CE Contingent Event

CoLL Cost of Lost Load

DER Distributed Energy Resources

EA Electricty Authority

ECE Extended Contingent Event

EFR Enhanced Frequency Response

EIPC Electricity Industry Participation Code

ER Extended Reserves

ERCOT Electric Reliability Council of Texas

EV Electric Vehicle

FFR Fast Frequency Response

FIR Fast Instantaneous Reserve

FK Frequency Keeping

HVDC High Voltage Direct Current

IL Interruptible Load

IR Instantaneous Reserve

KKT Karush-Kuhn-Tucker

LP Linear Programming

MILP Mixed Integer Linear Programming

MIQCP Mixed Integer Quadratically Constrained Programming

NFR Net Free Reserve

NR Negative Reserve

PFR Primary Frequency Response

PLSR Partially Loaded Spinning Reserve

QCP Quadratically Constrained Programming

RMT Reserve Management Tool

RoCoF Rate of Change of Frequency

SIR Sustained Instantaneous Reserve

SPD Scheduling Pricing and Dispatch

SR Spinning Reserve

TWDR Tail Water Depressed Reserve

VRE Variable Renewable Energy

LIST OF SYMBOLS

Offer Information

p_i and $P_i(t, p_i)$	IL dispatched from offer, i, and the response
p_i^{max}	Maximum IL offered
u_i and $U_i(t, u_i)$	SR dispatched from offer, i, and the response
u_i^{max}	Maximum SR offered
g_i	Ramp rate of SR offer
$t_{i,p}$	Time IL is initiated
$t_{i,u}$	Time SR starts ramping
$t_{i,e}$	Time SR finishes ramping
$t_{i,e}^{max}$	Maximum possible time SR can finish ramping
$c_{i,p}$	Price of an IL offer
$c_{i,u}$	Price of a SR offer

Frequency and Inertia

f	Grid frequency
f_{lim}	Minimum limit on grid frequency after contingency
f_j	Levels of the grid frequency limit
$t_{j,l}$	Time of transitions from one frequency level to another
t_{min}	The first time the grid frequency reaches its minimum
R	Risk, size of the largest credible contingency
P_M	Total mechanical power imparted through the turbines of synchronous generators
P_E	Total electrical power drawn from synchronous generators
H	Inertial constant

Counts and Totals

Nc	The number of transitions in the grid frequency limit
Ne	The number of equations in the standard form of the QCP problem
Np	The number of IL offers
Nu	The number of SR offers
Nr	The total number of regions

Ns	The number of slack variables

Convexity

q_i	Amount of reserve from an IL offer upon transformation
w_i	Amount of reserve from a SR offer upon transformation
$w_{i,j}$	Maximum amount of reserve that can be dispatched before $t_{j,l}$ for a dispatch u_i
$w_{i,m}$	Maximum amount of reserve that can be dispatched before t_{min} for a dispatch u_i
$t_{min,0}$	The earliest possible time the minimum frequency could occur
$A(\bullet)$	The amount of energy lost by the contingency and not replaced by time t_{min}
$F_j(\bullet)$	Amount of energy above what is required to keep the frequency above its limit by time $t_{j,l}$
$F_m(\bullet)$	Equivalent to $-2Hf_{lim} - A(\bullet)$ with a similar meaning to $F_j(\bullet)$
S_0	Set of all feasible points bound by the offer constraints and reserve requirement
S_j	Set of all feasible points to the frequency limit at $t_{j,l}$
S_F	Set of all feasible points satisfying all constraints
$T_j(\bullet)$	Non-linear vector transformation defined for each $t_{j,l}$
$T_m(\bullet)$	Non-linear vector transformation for the minimum frequency constraint
$\mathbf{v}$	Element of the problem space
$\mathbf{w}$	Transformed element from $\mathbf{v}$ in the same problem space

Region Definitions

p_m	Critical IL output to finally satisfy the reserve requirement
w_i^{min}	General minimum limit for a SR offer
w_i^{max}	General maximum limit for a SR offer
t_{min}^{min}	Minimum possible time t_{min} can be for a SR region
t_{min}^{max}	Maximum possible time t_{min} can be for a SR region
Q	Set of all IL offers
Q_B	Set of IL offers initiated before t_{min}
Q_E	Set of IL offers initiated at t_{min}
$Q_{B,j}$	Set of IL offers initiated before $t_{j,l}$
W	Set of all SR offers
W_B	Set of SR offers that start and finish ramping before t_{min}
W_T	Set of SR offers that start ramping before t_{min} but finish after
$W_{B,j}$ and $W_{T,j}$	Similar to W_B and W_T except with $t_{j,l}$

Chapter 1

INTRODUCTION

Variable Renewable Energy (VRE), such as wind and solar energy, is known to have a negative impact on power system security by reducing inertia on the power system, uncertainty in power output, and causing system oscillations [Shair et al. 2019], etc. Grid codes have improved the facilitation of VRE and their penetration has increased as concerns over climate change mount. The consequences of these issues were realised in the South Australian Blackout of September 2016 [AEMO 2017a]: the State of South Australia experienced a severe storm, lost connections to the rest of Eastern Australia, and wind turbines then disconnected from the grid after multiple voltage dips. The ensuing loss of generation in South Australia caused the remaining transmission circuit to the rest of Eastern Australia to trip on overload. With no available sources of generation to replace lost wind turbines, not even the emergency load shedding systems could halt the decline in frequency and complete collapse.

The outcome to this event has been a much publicised installation of a 100 MW h battery system at the Hornsdale Wind Farm (`hornsdalepowerreserve.com.au`), currently the largest lithium-ion battery bank in the world. Since then another 50 MW h and 30 MW h have been added in Victoria. This development is only the tip of the proverbial iceberg for new technologies: Quebec requires wind turbines to provide synthetic inertia [Asmine et al. 2018]; Great Britain [NG 2016], Ireland [EirGrid et al. 2014] and Australia [AEMO 2018b] [AEMO 2017c] are separately designing faster reserve categories; and North America issued a directive to remove barriers of entry to energy storage [FERC 2018].

New Zealand has decided to delay development of reserve markets, finding insufficient reason to adopt new technologies. Work on developing both Frequency Keeping (FK) [EA 2017b] and Instantaneous Reserve (IR) [EA 2018d] markets has stopped, and a report on battery technologies [Transpower 2017a] has found insufficient benefit for utility scale batteries to outweigh its cost. Current areas of work are normal frequency management [EA 2018c], the fair distribution of balancing requirements to generator governors; and emergency reserves [EA 2017a], i.e. AUFLS or now called Extended Reserves (ER). These are not likely to require new technologies. However, this has

not stopped batteries coming to New Zealand, as Vector (the electricity distributor for Auckland) installed a 2.3 MW h battery unit in the Auckland suburb of Glenn Innes.

The demand for reserves and their total cost in NZ over the last five years has declined to such a point that any development is considered to have limited benefit. In 2012, the total yearly cost of IR and FK was around $95 million, but in 2016 it was closer to $23 million. The reason for this is the completion of Pole 3 and upgraded controls of Pole 2, the two halves of the HVDC link between the North and South Islands. The new controls provided greater functionality, allowing for wider sharing of power between islands, so that the two synchronous networks act as if they have one frequency [Teeuwsen et al. 2013] [Phethean et al. 2015b].

NEW ZEALAND ELECTRICITY INDUSTRY

The New Zealand electricity industry is dominated by hydroelectric generation. Out of the 43.0 TW h of New Zealand's annual electricity demand, 26.1 TW h is provided by hydro generation, 60.6%. A further 23.3% of renewable energy is supplied by geothermal, wind, biogas, and wood resources; geothermal energy that comes from the central North Island forms the major share. Therefore New Zealand produced 83.9% of its electrical energy from renewable energy resources in 2018. The remaining 16.1% is produced from gas, coal, and oil entirely from the North Island. This information is obtained from the Ministry of Business, Innovation, and Employment (MBIE) electricity statistics.

If the proposed Climate Change Response (Zero Carbon) Amendment Bill [New Zealand Government 2019] is passed as law then net greenhouse gas emissions, other than biogenic methane, has to be zero by 2050. A clear step in achieving this target is obtaining a 100% renewable power system by replacing current thermal power stations with renewable ones, as stated by a report from Vivid Economics [Kazaglis et al. 2017]. Not just to replace the current 6.9 TW h of thermal electricity generation, but potentially the electrification of industrial processes and transport in general, i.e. the process of de-carbonisation. Wind energy will play an important part in meeting this target, as appropriate hydro and geothermal resources will diminish. Although this is a difficult statement to justify, its validity can be seen in what generation types have the most consents for construction. The Electricity Authority, the electricity industry regulatory in New Zealand, provides a list of consented generation on its Electricity Market Information website. Wind energy has 2517 MW of consented generation, whereas hydro and geothermal have 251 MW and 303 MW respectively.

Without anticipating future increases in electricity demand as a result of de-carbonisation, approximately 3000 to 4000 MW of extra wind generation capacity above the current 700 MW is required to replace 6.9 TW h of thermal electricity generation. The exact capacity required to achieve a 100% renewable power system is dependent on hydrology and wind resources for the year. This raises an important point about how

to define a 100% renewable power system, e.g. one possibility is that it means for a median hydrological year, electricity generation has to be entirely renewable.

New Zealand is blessed with renewable energy storage in the form of storage lakes, the greater quantity of which is found in the South Island. Lake Taupo and Lake Waikaremoana are the main storage lakes in the North Island, they can store 740 GW h together. New Zealand's total lake storage capacity varies between 3.5 to 4.0 TW h, depending on the allowable lake levels, which are based on the resource consents for each season. This amount of storage is quite small compared to the total yearly electricity demand of 43.0 TW h, especially after considering that national peak electricity demand is in winter for heating loads, and greater water inflows occur in the summer months as snow melt replenishes South Island lakes. Therefore there is already a significant challenge in managing New Zealand's energy resources. If wind energy is to replace thermal generation in supplementing hydro, then the uncertainty of wind generation in energy output is going to increase the challenge, as wind generation is likely to be less in the lead up and into winter, from April to July [Bull 2010]. The process of removing the last 16.1% of thermal generation will not come without its challenges.

The challenge this thesis focuses on is the real-time requirement for both generation and reserve capacity to meet both electricity demand and the reserve requirement from the largest credible contingency. This thesis particularly focuses on the reserve requirement. Wind energy is dependent on wind speed to determine its real power output, it cannot always be relied upon to meet the real-time electricity demand. Therefore some other form of generation capacity, or Demand Side Management (DSM), is required for these low wind speed conditions. Wind generation is also likely to increase the largest credible contingency in New Zealand, by increasing the peak power flow between the North and South Islands [Schipper et al. 2019]. Secondly, wind generation will decrease the inertia of the power system, and increase the requirement for faster acting reserves. It is difficult to separate the capacity adequacy problem into energy and reserves, as generation capacity can be divided between both. However, focus is placed on the reserve requirement with minimal consideration given to the energy side.

The New Zealand power system consists of two synchronous grids, one for the North Island and one for the South. They are connected via the HVDC link, which has a physical capacity of 1200 MW, but it is rarely required to surpass 1000 MW. This link is a critical component and requires the real-time coordination of reserves in the case it has an emergency. The North Island power system, with an electricity demand in the range of 1600-4600 MW, is larger than the South Island, with 1000-2200 MW. The total New Zealand electricity demand is in the range of 2700-6700 MW.

If 4000 MW of new wind generation is added to the current 700 MW, then in terms of the numbers, it is possible for New Zealand's electricity demand to be instantaneously balanced by wind generation, as 4700 MW of wind capacity is greater than the minimum

electricity demand of 2700 MW. However, for the security of the power system this will never occur, and some wind generation will have to be curtailed. This thesis analyses situations of increasing wind penetration, in scenarios adding 500 MW, 1000 MW, ... 4000 MW of wind capacity, and analyses the impact this wind capacity has on reserve requirements. To model curtailed wind generation, a principle is established of keeping all energy from current renewable generation. To determine the reserve requirement from contingencies, the largest contingency is monitored and the total inertia on the power system is modelled.

Instantaneous Reserve (IR), for which contingencies determine the requirement, are not the only reserve, and contingencies are not the only cause. Analysis has shown that wind generation will mostly affect the requirement for IR [Schipper et al. 2019], and consequently will be the most constrained aspect of reserve provision. This does not say that the requirements for other reserves will not increase as well.

RESERVE OPTIMISATION

Results show that if wind generation were to be built in the South Island to replace thermal generation in the North Island, then the increased power transfer north across the HVDC link would increase the requirement for IR in the North Island, as the loss of a single converter pole would be the critical contingency. Coupled with decreasing inertia in the North Island, the demand for faster reserve products becomes evident. Therefore New Zealand, like Australia, Ireland, and Great Britain, should consider whether options of new reserve technologies will be beneficial.

The means of determining the requirement for IR in the real-time co-optimised energy and reserves market of New Zealand suffers from a drawback. In situations of fast declining frequency for a possible contingency, reserve has to respond quickly, and to ensure this response speed, the IR requirement is made significantly larger than the contingency size. This is to guarantee the availability of more units. However, it may come at the cost of under-utilising generation capacity, which may be desirable for satisfying the electricity demand. To avoid this potential inefficiency, a method of including response characteristics of reserve providers into the optimisation is of benefit.

Another issue arises, how is it possible to know when new reserve technologies will be of benefit? It is possible to develop a new reserve category and create a market for it, like in Ireland and Great Britain. However, the requirement for these new reserves is dependent on assumptions about the rest of the power system, such as expected inertia, size of contingencies, and availability of conventional reserves. For example, may be it is better to build synchronous condensers to increase inertia, and to retain current reserve providers, than it is to build utility scale battery systems for fast reserve response. A methodology of making comparisons between options is required. Ultimately, the final solution will be predicated on robust dynamic simulations of keeping the grid frequency

above limits. Putting aside linking short-term scheduling with long-term planning, can these dynamic simulations be approximated in an optimisation formulation, which can consider options of changing inertia, changing contingency size, and the choice of reserve provider? The goal of this thesis is to show that this is possible.

The goal of this thesis is to construct an optimisation problem of the following form:

minimise *reserve availability cost*
subject to *frequency constraints*
and reserve availability limits

The objective is to minimise a linear function of the amount of reserve made available. It is possible to include inertia and contingency size into the objective, but typically these quantities are determined by energy constraints in co-optimised problems, and have been omitted. The reserve options have dynamic characteristics, starting with a delay, and then either an instantaneous or ramped response. The dispatched power of each reserve provider is determined by an optimisation whose objective function is to minimise the cost of reserve. Although reserve power can be stepped or ramped, while power output changes over time, it is the maximum power that is optimised. The choice of reserve has to create a frequency transient that is above the frequency limits. Lastly, limits are placed on final reserve output to recognise the finite capacity of reserve providers.

This optimisation problem is not linear, but quadratic. The constraints cannot be expressed in a closed form, but are piecewise. The feasible solution space is convex and the optimal solution can be found easily, as a fast solving methodology is created for it. Therefore the optimisation is still practical.

The full application of this implementation is to assess decisions that ensure frequency stability during contingencies. It has the potential of being applied in real-time electricity markets, thereby improving short-term efficiency. This optimisation can also be applied in longer duration markets as well, such as Day Ahead markets or Capacity markets. Therefore to retain generality of application, the optimisation is agnostic in terms of energy constraints. A full implementation into energy markets will require further research of how to include mixed integer variables into the solving methodology, as these are common in electricity markets.

MAIN CONTRIBUTIONS

The main contributions of this thesis to the literature are:

1. An analysis of the New Zealand power system under higher penetrations of wind generation, focusing on changes to reserve requirements. Integration studies of the New Zealand system have been completed before, but this research quantifies the likelihood of possible operating conditions, and to a greater level of wind penetration.
2. The construction of a fast means of optimising reserves for contingencies while taking account of speed of response and frequency limits. Included in this is
   (a) A proof of convexity for the feasible solution space.
   (b) A solving methodology for problems that are simultaneously both quadratic and piecewise in their constraints.

STRUCTURE OF THESIS

This thesis is divided into three main chapters. Before them is a short background. The first main chapter quantifies the amount of reserve required under higher penetrations of wind generation in New Zealand, to determine whether there is any motivation to develop reserve markets in New Zealand. This is Chapter 3, and is extracted from the report produced for the GREEN Grid project [Schipper et al. 2019]. The second main chapter, numbered 4, reviews methods in the literature of incorporating frequency constraints into MILP optimisations, which can generally be applied into real-time electricity markets. However, some optimisations are particularly applied to the Unit Commitment problem. The third main chapter, which forms the core of this thesis, develops a primary reserve optimisation formulation so that different providers, with varying response times, can be selected according to their performance. The formulation itself is developed in Chapter 5, but relies upon results obtained from the Appendices. Lastly, a list of future research is given in Chapter 6.

This thesis assumes a familiarity on the part of the reader to power systems, markets, and optimisation theory. A background is given in Chapter 2 to reiterate the main ideas. Chapter 5 assumes knowledge of Karush-Kuhn-Tucker conditions, particularly their geometric significance.

LIST OF PUBLICATIONS

By mid June 2019, three manuscripts had been produced awaiting review and publication: one report and two journal articles. The titles of these manuscripts, starting with the report, are:

- *Recommendations for Ancillary Service Markets under High Penetrations of Wind Generation in New Zealand*
- *Optimizing Instantaneous and Ramping Reserves with Different Response Speeds for Contingencies—Part I: Methodology*
- *Optimizing Instantaneous and Ramping Reserves with Different Response Speeds for Contingencies—Part II: Implications*

Chapter 2

BACKGROUND

The frequency of an AC power system is a critical variable. Improper frequency control will lead to the collapse of an entire AC network, and hours or even days before the grid can be electrified again. Within that time period the economic impact will be large, as commercial and industrial businesses cannot operate. The personal and social cost of a blackout is difficult to quantify, but is significant nonetheless. Therefore it is of absolute importance that this should not happen, and in NZ it does not happen often, because the power system has been properly managed. However, with increased penetrations of Variable Renewable Energy (VRE), new challenges are being encountered in securing the power system, as these new resources are fundamentally different machines to their conventional predecessors. This background gives the necessary information to appreciate the challenges, and provides context to understand the motivations developed in the next two chapters.

2.1 FREQUENCY CONTROL

Frequency characterises how quickly voltage on the grid oscillates between positive and negative. A voltage can be developed on the grid by a number of means, but the conventional way is by a synchronous generator. The synchronous generator has a rotor, spinning by the force of an energy source, with a constant magnetic flux. This flux induces a voltage via electromagnetic induction onto stator windings that are connected to the grid. The rate which voltage oscillates is proportional to the rotor speed, therefore controlling frequency is a matter of adjusting the speed of the rotor.

The frequency across a whole power system is the same everywhere. Although there is more than one synchronous generator on the grid, all their speeds are locked together by the effect of synchronising torque. If two synchronous generators deviate, equal and opposite torques are applied to the rotors to bring them back together. Therefore in its simplest approximation, all the synchronous generators on the power system can be modelled as a combined unit with a single frequency. If the exact frequency is measured at two different generators, then there will be little difference, because there are small

changes in rotor angle, the derivative of which cause small differences in frequency, that occur as generators retain synchronism.

Frequency changes depending on the balance of power coming in and out of the system. Power is conventionally added by sources of energy through a turbine, whether this is from hydro, steam, or gas turbines; this total power is referred to as the mechanical power. Power is transformed from mechanical to electrical by generators which feed loads. When there is an imbalance between mechanical power supplied through turbine and electrical power drawn by the generator, by the conservation of energy, the balance has to be satisfied by the rotational energy stored in the spinning rotors. If mechanical power is greater than electrical power, then the speed increases as the stored rotational energy rises. The sensitivity of the speed to variations in the balance of power is determined by the inertia. Greater inertia requires greater effort to change its speed, and results in more time available to react.

In New Zealand, frequency is controlled to maintain a near constant value of 50 Hz. The design of equipment is optimised for this speed and any major deviation from it will result in equipment damage. If frequency was allowed to drop below 47.5 Hz then Combined Cycle Gas Turbines (CCGTs) and other generation types [Brown 1998] [Ross 2001], will have to disconnect, as they start having problems with resonances. If more generation is lost by this point in time, then the frequency continues to fall and results in all generation being lost. Disturbances that result in a large frequency deviations, may not be root-causes of blackouts, but these disturbances do occur in the progress towards a blackout [Pourbeik et al. 2006]. Therefore every imbalance in mechanical and electrical power has to be managed to avoid continuing the frequency deviation. Control is primarily provided by power stations that can adjust their mechanical input, such as hydro [IEEE PES 2011].

In managing a power system, frequency stability is the ability of a power system to recover from a disturbance, so that frequency is brought back to a normal state. For frequency stability, these disturbances involve changes in real power from generation or demand [CIGRE 1999]. Frequency stability is one of three main categories for power system stability, the other two being rotor angle stability and voltage stability [Kundur et al. 2004]. The first is concerned about the synchronism of generators, and the avoidance of pole slip. Ensuring rotor angle stability for large disturbances, where linearising techniques are insufficient to model the situation, is called transient stability. Voltage stability is the ability of the power system to retain steady voltage within stable levels after a disturbance. Disturbances could be the result of changes in network topology, condition, or power flows.

Analysing power systems is necessary to anticipate issues and form mitigation measures to stability problems. This requires creating models of the power system, which include models of governor and excitation systems [Koritarov et al. 2013b],

turbines and generators, transmission systems, and loads [Kundur et al. 1994] [Anderson and Fouad 2003] [Krause et al. 2013] [Akhmatov 2003]. Detailed models are simulated on computers through commercial software packages, such as DIgSILENT PowerFactory, Powertech Labs DSATools, ETAP, etc.

2.2 RESERVES

The main balancing between generation and demand, i.e. between mechanical and electrical power, is done through the dispatch process. A dispatch is sent every five minutes in New Zealand for generators to adjust their set points, their base level of power output. The dispatch predicts future changes in demand as it fluctuates during the day, week, and year; working with longer term forecasts so that generator availability can be scheduled. However, there are events that cannot be predicted, and require capacity to be held available at short notice; this is called reserve. These events could be the unexpected loss of generation, the loss of transmission elements, variations in load, or error in predicting demand. Reserve is divided into three categories depending on the size and duration of the events it is managing. Following the classification of frequency control reserves in [Rebours 2008] and [Ellison et al. 2012], the categories are primary, secondary, and tertiary frequency control reserves.

- Primary reserve stops frequency deviating outside the allowable range, and manages real-time balancing. Primary reserve physically operates to manage imbalances that occur instantly or over a time period of about 2 minutes. However the exact definition depends on the power system. Some power systems have multiple subcategories of primary reserve, the names of which can use 'primary', 'secondary', and 'tertiary', but these should not be confused with broader categories used here. [Roberts 2018] provides a review of requirements for primary reserve provision across multiple power systems.

- Secondary reserve controls the frequency so that it is restored to nominal, 50 Hz in New Zealand, for large imbalances that have been managed but result in a steady state frequency deviation or for normal variations in load and generation. Secondary reserve operates to remove imbalances that remain after 1 to 10 minutes, or up to the time of the next dispatch. In its physical implementation, a centralised signal is sent from the system operator to specific generators to adjust their power output. Terms that also refer to this form of reserve include Frequency Regulation, and an application of Automatic Generation Control (AGC). If AGC is also used to control power flows amongst different areas of the network as well as controlling frequency, then it is called Load Frequency Control (LFC). In New Zealand, secondary reserve is called Frequency Keeping [EA 2016]. In Australia it is called Regulation FCAS (Frequency Control Ancillary Service) [AEMO 2015], for which

they have a causer-pays methodology to recuperate costs [AEMO 2017b], and is one the most technical methodologies for cost allocation.

- Tertiary reserve is used to relieve primary and secondary reserve if they are against their limits, and for large and slow events. This reserve is initiated through the generally manual actions of the system operator. Tertiary reserve providers have to be prepared to respond in the 5 minutes to several hours time frame. Tertiary reserve is also procured from power systems that have a significant amount generation that requires a warning time for it to come on-line. However, power systems like New Zealand with a large amount hydro generation that can respond at short notice, there is no market for tertiary reserve. With increased penetrations of VRE in international electricity markets, tertiary reserves is being procured to ensure there is enough generation with sufficient ramp-rate [Kirby 2014], so that the daily demand changes can be followed, and the changes in VRE output over a day.

There is also another form of reserve, emergency reserve, used for events so large and fast, primary reserve is not able to arrest the decline in frequency. Emergency reserve is provided by quickly disconnecting loads, thereby sacrificing some customers electricity supply temporarily in order to retain the stability of the power system that remains. In New Zealand, three events of this magnitude have caused the operation of emergency reserves in the last decade: one each in 2011, 2013, and 2017. Loads are placed into blocks, so that when a certain frequency is reached a block is disconnected [Young 2009]. New Zealand is going through a process of improving emergency reserve operations, while at the same time improving the discrimination between important and non-critical loads [Transpower 2017b].

This thesis is interested in primary reserve, as it is the most affected by increased VRE penetrations in New Zealand. Although, this does not mean the requirements for other reserve types will not increase. This is further explained in the next chapter.

2.3 COST EFFECTIVE CONTINGENCY MANAGEMENT

There are three main costs to consider when managing contingencies that give rise to the demand for primary reserve. Each of these costs are connected and require trade-offs to find the optimal design. They are:

1. The cost of extending the capabilities of equipment. In New Zealand during 2001 and early 2002 [FSWG 2001], frequency standards were reassessed in the North Island, so that generators were allowed to disconnect instantly if the frequency were to drop below 47 Hz. Previously, generators were required to remain connected down to 45 Hz for a short period. This restricted what types of generation could be built in the North Island. Therefore instead of meeting the cost of extended capabilities of potentially new generators, the frequency standards were raised. The frequency standards for each country can be found in their grid codes: New Zealand [EA 2018b], Australia [AEMC 2017], Ireland [EirGrid 2015].

2. The cost of unserved electricity to customers (Cost of Lost Load (CoLL)). Electricity provides substantial benefit to industrial and commercial customers, and to personal well-being, if that electricity were no longer available then it will come with significant cost. Therefore in making decisions that determine the reliability of a power system, it is necessary to quantify these costs. In New Zealand, estimates are provided in [CAE 1993]. However, recent work on Extended Reserves in New Zealand provides the latest values [NZX 2017].

3. The cost of primary reserve. There is a cost in providing reserve, i.e. having the necessary equipment and maintenance, and an opportunity cost if the same capacity were provided for energy production.

In any given power system, there will be a number of power stations. There is a certain probability at some point in time that one of these generators will disconnect in an emergency, therefore enough reserve has to be held to cover this contingency (often called risk in New Zealand). It is theoretically possible that all power stations could have an emergency at the same time, but the likelihood of this is so low, it becomes impractical to have a double set of power stations held as reserve to cover all current generation. Therefore some contingencies cannot be covered, as they are too large and unlikely, so a definition of credible event is required to determine which contingencies are valid concerns, such as the single loss of generation.

Secondly, reserve requires a finite amount of time to respond before frequency falls below its limit. If there is insufficient time then faster reserve is required at greater cost. There is another option, the capabilities of equipment can be extended to have a larger tolerable frequency range.

The problem of determining the optimal trade-off between equipment capabilities, definition of credible event, and reserve provision is clearly difficult. To explicitly evaluate all these costs over a range of possible conditions is not practical, therefore the usually approach is to set a series of policies. Equipment capabilities are standardised in codes, these standards are largely developed from capabilities of existing equipment as they are difficult to change once they are already in operation.

The level of risk that has to covered by reserve is determined by a minimum likelihood of contingency [Transpower 2014]. If an event is less likely to occur than this minimum, then reserve is not procured for that contingency. However if it is more likely, then reserve is procured. In New Zealand, this is formed as a $N-1$ policy [Transpower 2017c], the redundancy level is set so that if one piece of equipment is lost, then there should be no loss of connection to customers. Similar policies exist in North America [NERC 2018].

2.4 ELECTRICITY MARKETS

In New Zealand, primary reserve for contingencies is provided by generation and disconnecting non-critical loads in the Instantaneous Reserve (IR) market. Other primary reserve for normal balancing is mandated through the Electricity Industry Participation Code (EIPC). Since generation capacity has the primary purpose to provide energy, reserves are co-optimised with energy, to find the optimal use of generation capacity. If the largest generator is the cause for the largest credible contingency, then it is also important to optimise this value as well.

New Zealand's real-time electricity market, mediated through the optimisation process, provides a competitive means of procuring reserve. Competition is hoped to bring efficiency, not just in the short term but more importantly in the long term, as the main cost of reserve is from the initial fixed cost of capacity. For IL this is less applicable as they have unserved energy costs, a short term cost, but the loads chosen for this service are deferrable, and contingencies that require IL are uncommon, so this cost should be competitive.

The electricity market should provide the right incentives for investors, signalling the right time for new capacity to be built. However the ability for the market to do this has been a matter of debate [Evans 2017], especially when incorporating new generation sources like wind and solar [Philpott et al. 2018]. There is also debate on whether markets are competitive [Wolak 2009] [Poletti 2018], but this problem is of less concern than avoiding energy shortages [Evans and Meade 2005], which receives more attention from regulatory bodies. A part of guaranteeing energy supply means having enough installed generation capacity. The scarcity pricing mechanism is applied in New Zealand to further incentivise capacity investment, but overseas networks have implemented capacity markets [Cramton and Ockenfels 2011].

Electricity markets in general require an optimisation formulation to be created. Formulations consist of an objective function, where cost is minimised, and a set of constraints that restricts the feasible solution space. Constraints model the limitations that exist in the market, e.g. the direction of power flow through transmission lines or the maximum power output from generators. Formulations are classified into different problem types depending on the nature of the objective function and constraints. The simplest form with constraints is Linear Programming (LP), with a linear objective function and linear equality and inequality constraints. Solvers are developed for these problems, such as Simplex or Interior Point methods. [Chong and Zak 2013] and [Wright 1997] provide introductions to these types of solvers. These solvers are found in commercial software such as CPLEX, and application based software such as AIMMS.

New Zealand's electricity market formulation can be found in the Scheduling Pricing and Dispatch (SPD) tool documentation [Transpower 2018c]. Before IR sharing through the HVDC link, the New Zealand formulation in its simplest form was a LP problem. However physical constraints are usually not linear, such as the flow of electricity through a network, therefore linearisation techniques are required. A DC power flow equivalent model can approximate real power flows on an AC networks [Schweppe 1988]. Variables that can only be discrete values are common, if they are bound by linear equations, then these problems are called MILP. In New Zealand, these are sometimes used to remove non-physical power losses from the HVDC link, and in determining reserve sharing limits between the North and South Islands.

Electricity markets are one application of power system optimisation, particularly towards competitive procurement. However, optimisations can be performed on power systems by a centralised coordinator. There exist categories for power system optimisation types. These include Economic Dispatch (ED), Unit Commitment (UC), Optimal Power Flow (OPF), and the inclusion of Security Constraints (SC), etc. These categories all minimise operational costs, but differ in costs and constraints. Unit Commitment seeks to optimise both fuel costs and the cost of starting and stopping generators, whereas Economic Dispatch is directed towards fuel costs. Optimal Power Flow considers power flow limits, voltage limits, and the cost of transporting electricity. Security Constraints recognise that a power system has to remain within operational limits if a contingency were to occur. These categories can be applied in electricity markets. Unit Commitment type problems apply well to Day-Ahead Markets, where a schedule is committed to on the previous day for the current 24 hours. Real-time markets that determine the final dispatch of resources use variations of OPF with the addition of SC.

The way reserves are included into the scheduling and dispatch process can be separate to the energy optimisation, or co-optimised with it. However, in either way, coordination between energy and reserves is required. This thesis develops a new reserves optimisation in Chapter 5, it assumes coordination with energy is required, but

is not explicit in its implementation. For this new approach to implementable in New Zealand, more research is recommended for joining the new approach to the current co-optimised electricity market.

2.5 VARIABLE RENEWABLE ENERGY

International concern over climate change has caused a global uptake in VRE, such as wind and solar energy. Hydropower, which is controllable provided it has sufficient storage, will also play a major part in mitigating greenhouse gas emissions, but there is not enough to replace all thermal generation. Therefore VRE has seen significant development in the last two decades. Uncertainty in power output is the main drawback of VRE, as wind speed and solar radiation vary with weather and time of day, power output fluctuates depending on the availability of these resources. VRE has three important implications for power system reserve requirements in New Zealand: a reduction in total grid inertia, minute to minute variability in power output, and the ability to change the size of the largest credible contingency. There are other issues but they have not been of direct concern in determining reserve requirements in New Zealand so far.

Type three and four wind turbines, which are the most common type [Yaramasu et al. 2015], and photovoltaic systems share an inverter based connection with the grid. This stops them experiencing any synchronising torque on their rotating components, if they have such a component, and therefore do not contribute or store any energy when there is an imbalance between mechanical and electrical power. For wind turbines this has the benefit of allowing it to optimise rotor speed to wind speed to extract the most energy out of the air. The downside is they do not intrinsically contribute in contingencies to slow the decline in frequency. They can provide a response to emulate synchronous machines, but this requires additional controls. In modelling the grid, this effect is seen as an absence of inertia from that generator, and if VRE replaces synchronous generation then the total inertia reduces. If the size of the largest credible contingency were to occur during these low inertia periods, then the requirement for primary reserve will increase.

The second implication is due to the uncertainty of wind speed and solar radiance on the minute to minute time scales. These variations have minimal impact, but require additional secondary reserve to sustain the same frequency quality. Minute to minute variations in VRE will utilise primary reserve more often that without it, but the capacity required to cover these variations is less than that of contingencies, except for possible scenarios where the largest credible contingency is small. Therefore VRE rarely has any impact on the amount of capacity required for primary reserve when managing uncertainty in power output.

Integration studies are conducted to determine allowable penetrations of VRE.

Several have been applied to New Zealand with varying degrees of detail [SO 2007] [Transpower 2017d]: McQueen [2016] provides a commentary on these.

Moving away from reserve requirements in New Zealand, issues experienced in other systems are briefly commented on. Frequency stability is a concern in international power systems. Many systems have recorded changes in total inertia. In Australia, historic inertia has been found for each state in Eastern Australia [AEMO 2018c]. Although mainland Eastern Australia is one synchronous network, with Tasmania asynchronously connected with an HVDC link, there is concern about events that will island parts of the network. South Australia is the most critical state with inertia dropping to a minimum of 1000 MW s in the 2016/17 analysis period. This is significantly less than the South Island of New Zealand which reaches a minimum inertia of 5000 MW s. This has prompted the building of synchronous condensers in South Australia [Appleby and Rositano 2019]. ERCOT with a minimum inertia of 130,000 MW s [ERCOT 2018]; Nordic system with a minimum of 120,000 MW s [ENTSOE 2017]; and Great Britain with a minimum inertia around 130,000 MW s [Ashton et al. 2015] [Palermo 2016], are all considering the impacts that this might have on frequency stability, transient stability, and Rate of Change of Frequency (RoCoF).

In Ireland, the recovery of real power output from modern wind turbines after a voltage dip is a concern to frequency stability [O'Sullivan et al. 2014]. Since modern wind turbines are given more time to recover to avoid mechanical damage, the reduction in real power output for an extended time could cause a frequency transient. Therefore a market is developed for the system service called Fast Post-Fault Active Power Recovery (FPFAPR) [SEM 2013] to incentivise faster responses.

Looking towards the future with power systems reaching 100% renewable energy from mostly converter based VRE, there are challenges. Retaining transient and frequency stability under very low inertia conditions will require new approaches to controlling converter based energy sources [Ackermann et al. 2017], in order to avoid curtailing VRE in these conditions. Curtailment can also be because there is too much generation to supply the electricity demand at any point in time. To mitigate this, energy storage can help [Kroposki et al. 2017]. However, curtailment already exists in countries like Great Britain and Germany, because of insufficient transmission capacity [Joos and Staffell 2018]. [Bird et al. 2016] reviewed VRE curtailment in 11 countries, finding that significant curtailment has been experienced in China, Italy, and the ERCOT system in some years as a result of transmission congestion.

For further information about broader integration issues, consult the textbook [Ackermann 2012].

2.6 NEW RESERVE PROVIDERS

The goal of reaching 100% renewable energy has motivated research into new technologies. A large focus has been on developing energy storage to counter-balance the uncertainty of VRE, but research has also included new forms of reserve, which are of interest to this work. These solutions are divided into three different categories: VRE, energy storage, and the demand side based approaches.

VRE, in its normal mode of operation, seeks to maximise energy output. This mode can be suspended for the provision of reserve, which can happen in two ways [Díaz-González et al. 2014] [Singarao and Rao 2016]: power output is limited below maximum, thereby providing enough headroom to offer reserves; or follow the normal mode of operation until a contingency occurs, and then utilise an intrinsic form of energy storage to offer a temporary boost in power output; this includes inertia emulation from wind turbines. The first has a clear disadvantage of having to spill a significant amount of energy, and only becomes viable if the reserve price is comparable to the energy price. The second spills a minimal amount of energy, but requires a recovery period to replenish the intrinsic energy source, therefore can only help in limiting the initial decline in frequency until other reserve is available.

HydroQuebéc has required inertia emulation from wind turbines since 2006 [Fischer et al. 2016]. Operational experience has been gained, and the impacts of recovery period on continued frequency stability was analysed. A prolonged recovery period was recommended [Engelken et al. 2017] [Asmine et al. 2018]. In New Zealand requirements for inertia emulation have been considered [Pelletier et al. 2012], but has not been implemented. A proof of concept to supply reserve into the IR market was demonstrated at Te Uku Wind Farm [Brown and Scott-Dye 2012] in New Zealand, this required operating below maximum output.

Energy storage comes in many forms [Aneke and Wang 2016]: pumped hydro, battery, flywheel [Mousavi G et al. 2017], compressed air [He and Wang 2018], super capacitors [González et al. 2016], and superconducting magnetics. Although pumped hydro is not considered a new technology, it is becoming more common for power electronics to be adapted to it so that energy output is maximised [Koritarov et al. 2013a]. Power electronics allow for fast reaction to grid conditions, mitigating the effects of declining grid inertia. Power electronics can improve efficiency of hydroelectric plants, allowing the turbine speed to be optimised against head and flow. The options for converter topology are similar to those for wind generation, with both doubly fed and full scale converters being options [Singh et al. 2014]. Pumped hydro energy storage is important in storing large quantities of energy [Geth et al. 2015] [Rehman et al. 2015] [Barbour et al. 2016] [Kong et al. 2017], and the mitigation VRE curtailment.

Battery Energy Storage System (BESS) are becoming prevalent for providing grid services. Battery types vary [Hu et al. 2017], but lithium-ion batteries are prevalent. In

Australia, batteries already provide reserves [AEMO 2018a], and in the United States they provide frequency regulation [Zhang et al. 2017] [Chen et al. 2017].

Demand side control is not a new technology, as Interruptible Load (IL) has been a part of New Zealand since the upgrade of the HVDC system in the early 1990's [Smith 1989], and is used in other countries as well. IL utilises deferrable loads in industry, such as pumps, motors, and large refrigerators. However more research has looked into using smaller commercial and residential loads that use a form of thermal energy storage [Dehghanpour and Afsharnia 2015], thereby being able to delay or bring forward power use in order to offer reserves. This has the benefit of using resources already available, but suffers from the cost of adding control circuits to each of these individual units. Electric Vehicles while connected to the grid, can through changing charging state provide reserves [Peng et al. 2017].

Chapter 3

RESERVE REQUIREMENTS

This chapter analyses the demand for Instantaneous Reserve (IR) under higher penetrations of wind generation in New Zealand. A set of scenarios is developed, starting with adding 500 MW, and incrementing by 500 MW until 4000 MW is added above the current installed wind generation of 690 MW. This is a total of eight scenarios. Progressing through each scenario, the issues that may arise as New Zealand tries to achieve 100% renewable energy are recognised. Wind energy has two implications on the requirement for IR, the reduction in inertia and changes in risk. Therefore it is necessary to model the generation dispatch process under each scenario, to calculate the total inertia, to find the largest generator, and the transfer of electricity across the HVDC link between the North and South Islands.

In New Zealand, IR is used to ensure the grid frequency does not drop below 48 Hz for Contingent Events (CEs), events considered to occur often enough that it is uneconomical to utilise AUFLS in the frequency's recovery. These events include the loss of a single generator unit, or the loss a single pole in the HVDC link. There is also a class of Extended Contingent Event (ECE), where both IR and AUFLS recover the frequency, the most common of these events is the loss of both HVDC poles. IR follows the definition of primary reserve in Section 2.2.

The majority of this chapter is extracted from a report produced for the GREEN Grid project: *Recommendation for Ancillary Service Markets under High Penetrations of Wind Generation in New Zealand* [Schipper et al. 2019]. The scope of this report is wider than the focus of this chapter, as it considers the impacts of wind generation on other reserve types. Some of these, like governor action in normal conditions and FK, are expected to be used more. However the amount is not uncommon amongst previous operating conditions experienced in NZ, therefore it is within existing capabilities. The report also analysed reserve types that do not currently exist in New Zealand, specifically tertiary reserves. It found that given the existing capabilities of hydro power stations, and the removal of thermal generation under higher penetration, that demand for these products will be minimal. The report found that there are significant changes in IR requirements, thereby providing the focus of this research.

3.1 NEW ZEALAND FREQUENCY MANAGEMENT

Before considering IR in detail, a brief overview of frequency management in New Zealand is given. Frequency management starts with the regular dispatch of generation to satisfy the electricity demand. These dispatches usually occur every five minutes. To manage the power imbalances that occur within five minutes and the error in the dispatch, there are mandatory requirements for generators to have governor control systems. To supplement normal frequency management, a second service, called Frequency Keeping (FK), is competitively procured. 30 MW of FK is procured evenly across both Islands. The System Operator sends a regular signal to each FK provider every 4 seconds to adjust set point. FK is New Zealand's equivalent to Automatic Generation Control (AGC), although it does not consider area balancing requirements that comes with full Load Frequency Control (LFC). This is because New Zealand is a small network with only one System Operator, one transmission grid owner, i.e. Transpower, and one electricity market. FK is not co-optimised with energy and IR, but its optimisation is executed first, and provides inputs into the larger energy and reserves optimisation. This is done every 30 minutes.

Governor action from generators is mandated through the Electricity Industry Participation Code [EA 2018b], and is used to manage contingencies, such as the loss of a generator or HVDC pole. The mandatory requirements do not ensure that reserve providers have the available capacity, then IR is competitively procured to ensure this availability. This has allowed loads that can quickly disconnect from the power system to provide reserve and to be paid for that service, as well as for conventional generators to improve their response characteristics.

Exceptional events, where the frequency drops very quickly but do not occur often enough to warrant procuring addition reserve, have the last set of options of quickly disconnecting loads at the 33kV feeder level. This service has been historically called Automatic Under-Frequency Load Shedding (AUFLS), but recent development has changed the name to Extended Reserves. Approximately 32% of demand is organised to disconnect over a series of blocks.

Contingencies that involve the HVDC link, quickly reducing power transfer, can result in high frequencies in the sending Island. Therefore Over-Frequency Reserve is procured. This is particularly procured from South Island generators, as north transfer is the highest.

HVDC link also supplies frequency management services, by equalising the frequencies between the two Islands. This allows for better normal frequency management in both Islands, as random fluctuations in power imbalance in each Island can cancel each other out. Also, it allows for the sharing of IR between the two Islands. However, this is not helpful if the HVDC link is the cause of the contingency.

3.2 INSTANTANEOUS RESERVE

The New Zealand IR market, coupled with the energy market, is one of the most technical formulations in the world. Its inclusion of dynamic simulations to determine constraint parameters for each trading period, every 30 minutes, is likely to be unique amongst international markets. A simplified description is presented to help understand the limitations it has, and to justify the approximations used to model the dynamic simulations, which determine IR requirement.

The general constraints for IR in the market formulation are as follows:

$$\forall c \in CASE: \quad \sum_{i \in F_c} RES\_FIR_i + NFR_c \geq RISK_c \tag{3.1}$$

$$\forall c \in CASE: \quad \sum_{i \in S_c} RES\_SIR_i \geq RISK_c \tag{3.2}$$

$$\forall i \in G: \quad GEN_i + RES\_FIR_i \leq GEN\_MAX_i \tag{3.3}$$

$$\forall i \in G: \quad GEN_i + RES\_SIR_i \leq GEN\_MAX_i \tag{3.4}$$

$$\forall i \in G: \quad 0 \leq RES\_FIR_i \leq FIR\_MAX_i \tag{3.5}$$

$$\forall i \in G: \quad 0 \leq RES\_SIR_i \leq SIR\_MAX_i \tag{3.6}$$

IR comes in two forms, Fast Instantaneous Reserve (FIR) and Sustained Instantaneous Reserve (SIR): fast implying that dynamic properties are of interest, and sustained for static properties. The first to ensure the frequency limits are not violated, and the second so that any generation lost is entirely replaced and the frequency will return to 50 Hz. In the formulation, reserve availability is marked by the variables RES_FIR_i and RES_SIR_i respectively, and are indexed by i the provider of that reserve. The set G is the range of all possible providers.

The amount of reserve offered must be within the capabilities of the providers, Eq. 3.5 and 3.6. This brings into consideration the definition of FIR and SIR, which depends on the type of reserve. There are three types: Partially Loaded Spinning Reserve (PLSR), Tail Water Depressed Reserve (TWDR), and Interruptible Load (IL). PLSR is offered from power stations which are operating below their maximum output. TWDR is specific to hydro power stations that operate their units synchronised to the grid but with water depressed from the turbine chamber. PLSR and TWDR can be generalised as Spinning Reserve (SR) in terms of IR definitions. IL is provided by

deferrable loads that have automatic disconnection devices triggered when the frequency drops below 49.2 Hz. The definition of FIR in NZ is the amount of reserve available six seconds after the contingency starts for SR, and one second after the frequency drops below 49.2 Hz for IL. For SIR, the definition is based on a 60 second average output after the contingency. There is a slight difference between SR and IL, for the former the average starts from the contingency, and for the latter from when the frequency is 49.2 Hz.

Although FIR and SIR may be referred to as two different products, they are not mutually exclusive, as seen in Eq. 3.3 and 3.4. These equations ensure that capacity of a provider is not over supplied between energy, GEN_i, and reserves. It is noticed that the same capacity can provide both FIR and SIR.

Eq. 3.1 and 3.2 are reserve requirement constraints. The equation for SIR is self explanatory, as the total amount of SIR has to cover the largest credible contingency and in New Zealand is called the risk for short, $RISK_c$. For FIR a parameter, NFR_c, is added to the constraint to account for power system frequency dynamics. NFR stands for Net Free Reserve, which depending on conditions of the power system determines the inherent reserve available that is not reflected in the offers. This could be from generators that have governors but have not offered their response as reserve, or from the natural response of loads, and as shown later, from inertia. There are a series of these constraints for each potential situation, c; $CASE$ is the set of these contingencies. For example, if the largest risk in the South Island is the loss of a single 120 MW unit from Manapouri hydro power station, therefore reserve offered from the South Island has to cover this risk; $F_c \subset G$ and $S_c \subset G$ are sets of these South Island generators, not including Manapouri, for FIR and SIR categories respectively. Other situations include events in the North Island, and the HVDC link. The IR market has four prices, determined from the Lagrangian of these reserve requirement constraints, one for each combination of FIR or SIR and whether it is NI or SI.

The description of IR has been simplified; the exact formulation for the market optimisation has more components than those expressed in Eqs. 3.1 to 3.6. For further information about these details, the SPD formulation should be consulted [Transpower 2018c]. The full equation set allows for reserve sharing between islands, and necessary restrictions in HVDC operation. It should be recognised that these documents are living, and subject to revision and expansion as the market is developed, so earlier revisions should be consulted to understand previous operations. For more details about the definitions of reserves and other market operations in NZ attention is directed to the following documents:

- Electricity Industry Participation Code [EA 2018b]. Provides the primary definition of reserves. Also includes reserve cost allocation, equipment frequency capability requirements, the standard for frequency excursions that the System

Operator has to maintain, the requirements for generator governor action and stability, the market offer structure and scheduling time frames.

- Transpower Policy Statement [Transpower 2017c]. Classifies contingent events into separate categories, therefore determining the level of action required and frequency range available to satisfy the requirements of EIPC.
- Credible Event Review [Transpower 2014]. Reviews the likelihood of events and therefore adjusts the classifications of the Policy Statement.
- Ancillary Service Procurement Plan [EA 2016]. Arrangements for procuring Ancillary Services.
- Companion Guide for Testing of Assets [SO 2016]. Provides greater definition for reserves through setting the testing requirements of maximum reserve capabilities.
- SPD Schedule Inputs [Transpower 2018a]. Describes inputs for different schedules, dispatch, and pricing.

The rest of this section is focused on determining the Net Free Reserve (NFR_c) parameter through the Reserve Management Tool (RMT) [Transpower 2018b]. The purpose of IR is to stop the frequency from falling below 48 Hz for a CE. This is done by adjusting the NFR parameter so that a sufficient amount of FIR is dispatched. RMT iterates over several values of NFR_c, determines the reserve to dispatch, and then performs a dynamic simulation to see how close the minimum frequency reaches 48 Hz. If the minimum frequency is a above 48 Hz, then RMT increases NFR slightly on the next iteration, thereby reducing the amount of FIR required. If the minimum frequency is below 48 Hz, it reduces NFR, and increases the FIR requirement. Transpower provides an excellent animation of this process: `www.transpower.co.nz/system-operator/about-system-operation-service/learning-centre`. Once an appropriate value is found for NFR_c it is then sent to the market optimisation, i.e. the SPD formulation.

To describe the influence of inertia and risk on the amount of FIR required, it is necessary to explain the equations of the power system. The grid frequency is dependent on the balance between mechanical and electrical power, this is expressed by the swing equation:

$$2H\frac{df(t)}{dt} = \Delta P_M(t) - \Delta P_M(t) \tag{3.7}$$

where f is the frequency normalised by the nominal frequency of 50 Hz, often referred to as the per unit value, and is also a perturbation from 1 per unit, so that when the frequency is 50 Hz, $f(t) = 0$. The change in frequency is determined by the power balance between mechanical power imparted to the turbines of synchronous generators, $\Delta P_M(t)$, and $\Delta P_E(t)$ is the electrical power drawn from those same generators in MW.

H is the inertial constant in MW s. It is assumed that rotational energy provided by some loads is accounted for in this inertial constant. Frequency falls quickly if inertia is low and risk is high, and response speed has to be fast to avoid reaching its limit.

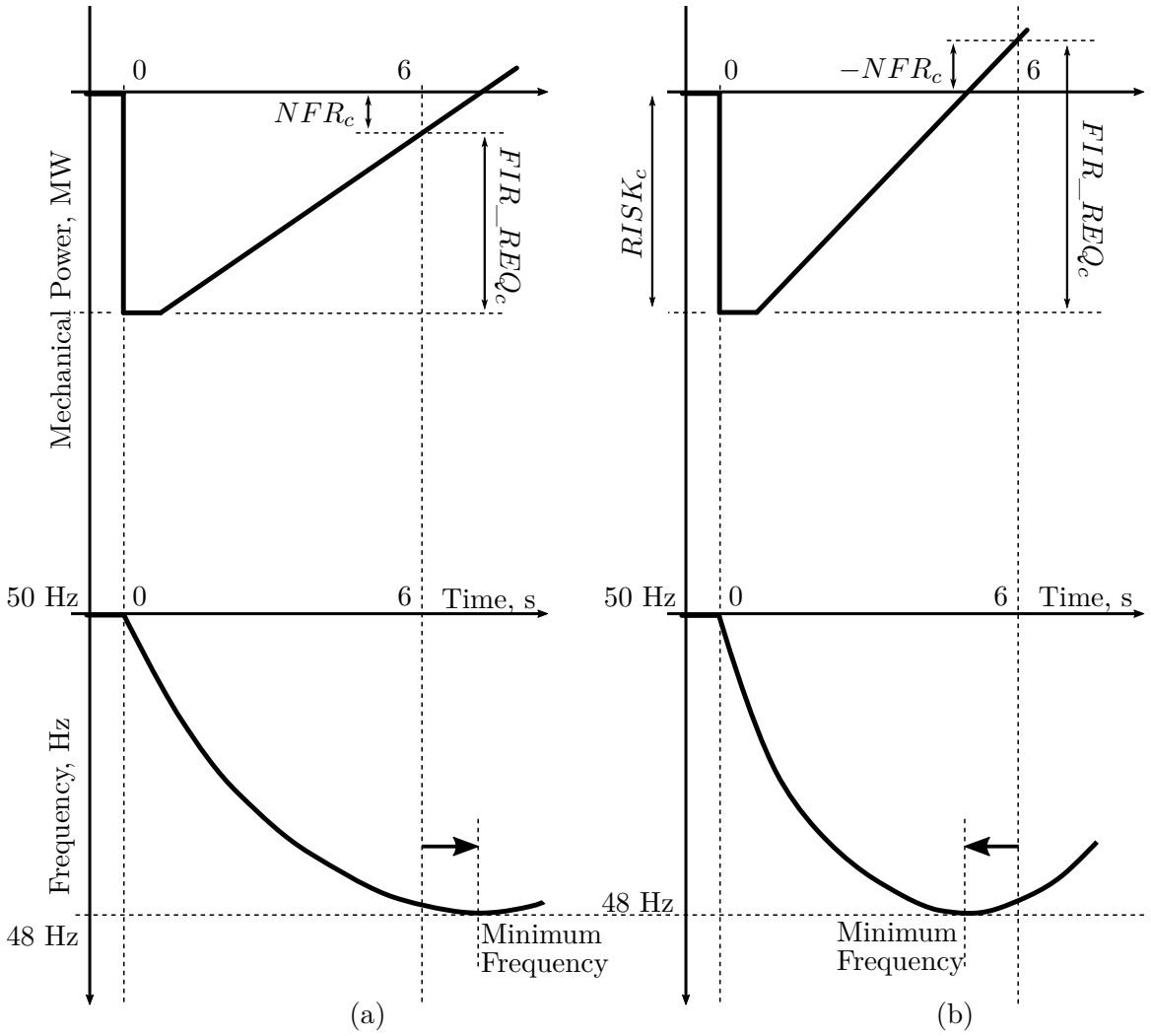


Figure 3.1 The required reserve response to keep the grid frequency above 48 Hz, for a high inertia case, (a), and a low inertia case, (b).

Consider Figure 3.1, which plots the frequency and the mechanical power supplied transients for two different values of the inertial constant in a loss of generator contingency; risk is kept constant. The first situation, (a), is with high inertia; enough reserve responds so that frequency reaches its minimum on the limit. The value of NFR_c obtained to achieve this result is sent to SPD as the requirement to keep the frequency above the minimal limit. Since inertia is high the rate at which reserve responds does not have to be fast, and the minimum frequency is reached after 6 seconds. This is seen as the right directing arrow of Figure 3.1a, showing the difference in time between 6 seconds and the minimum frequency. The amount of FIR required is then less than the

risk, as it is not necessary to require the full amount by 6 seconds in time. This is seen by the mechanical power balance not quite replacing the risk by the time of 6 seconds in Figure 3.1a, that is the 6 seconds for which FIR SR capability is defined. Therefore NFR_c is positive. To ensure the frequency does recover, SIR is procured to cover the risk.

For the second situation, (b) with low inertia, frequency falls quickly. The response from reserve has to be fast and the minimum frequency reaches the limit before 6 seconds, as shown by the left directing arrow of Figure 3.1b. The amount of FIR required is greater than the risk at 6 seconds, as ensuring this ramp rate can only be guaranteed if the amount of FIR procured implies it. Therefore NFR_c is negative. In an actual event, the full amount of FIR may not be utilised, as procuring reserve above the risk ensures enough units are not operating at their maximum output in order to sustain the required ramp rate. The response rate is important for these situations where the minimum frequency is reached before 6 seconds.

To show this in practice, Figure 3.2 shows the distributions of total FIR and SIR procured in the past for both Islands. Focus on the results prior to 2015, before reserve sharing increased between the two Islands. The amount of SIR is a good indication of the risk: for the North Island this is commonly around 400 MW, the amount of generation from either Otahuhu, Huntly, or Taranaki CCGT. For the South Island the largest contingency is usually 120 MW, the size of a single unit at Manapouri hydro power station. The amount of FIR reserve procured is noticeably less than SIR for both Islands by comparing the plots of Figure 3.2. Hence for most of the time inertia and risk are such that the first situation of Figure 3.1, (a), is most likely.

There is an exception found in first two quarters of 2012 in the South Island, where the amount of FIR procured is significantly more than SIR. During this time, South Island lake storage was low, requiring southward transfer of power across the HVDC link. This was before Pole 3 had been fully connected, and while the older Pole 1 was still in use. Up to this time for southwards transfers, the HVDC link lacked the capability of near instantaneously transferring power from one pole to the other in case the first had an emergency and needed to be disconnected. Therefore the South Island had to cover the risk of a single pole, roughly half the HVDC transfer, and considerably more than a single Manapouri unit. Inertia was also low, as hydro generation was conserving storage levels by not running, hence the system resembles situation (b). This is seen in Figure 3.2 by the requirement for FIR being greater than the requirement for SIR in first two quarters of 2012 in the South Island.

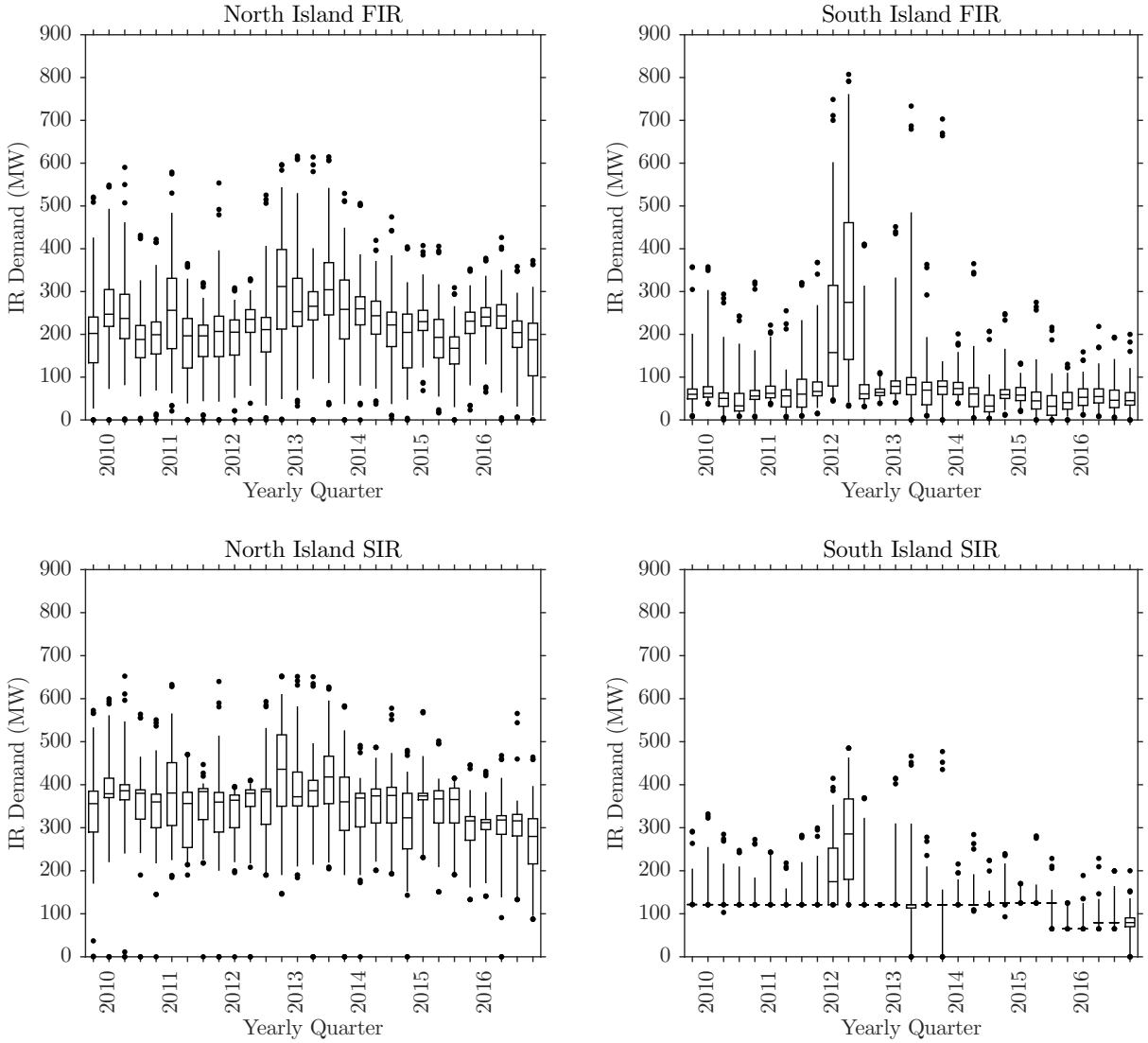


Figure 3.2 The distribution of demand for IR. The distribution is represented by a box and whisker plot. The top and bottom of the box are the 25^{th} and 75^{th} percentiles. The whiskers represent the 1^{st} and 99^{th} percentiles. The dots are the three maximum values and three minimum values. The time range of the data extends from the last quarter of 2009 to the last quarter of 2016.

3.3 FIR REQUIREMENT MODELLING

To quantify the impacts of each wind penetration scenario on the requirement for FIR, a model is produced to approximate RMT. Transpower originally calculated FIR requirement with a MathWorks Simulink model, but the dynamic simulations are now performed through Powertech Labs Transient Security Assessment Tool (TSAT). The models implemented in these tools are complex, and to reconstruct them is not the purpose of this research, so a simplified model is developed instead. Secondly, the information in these models is the property of generators, and requires obtaining agreements with each of them, so this data is difficult to access.

The first step in creating a simplified model for the requirement on FIR is to limit reserve provision SR and to omit IL. This is not to say that IL will be removed in the future, but it is easier to solve the equations with only one type of reserve present, and removes the need to estimate the relative proportions of SR and IL. Appendix A of [Schipper et al. 2019] provides an analysis of incorporating IL, and it should be understood that including IL will change the results obtained in this chapter, therefore the results should be interpreted in light of this simplification. The mechanical power, $\Delta P_M(t)$, is simplified to a form shown in Figure 3.3, and expressed as:

$$\Delta P_M(t) = \begin{cases} 0 & t < 0 \\ -P_r & 0 \leq t \leq \tau_d \\ -P_r + P_r \frac{t-\tau_d}{\tau_r} & t \geq \tau_d \end{cases} \tag{3.8}$$

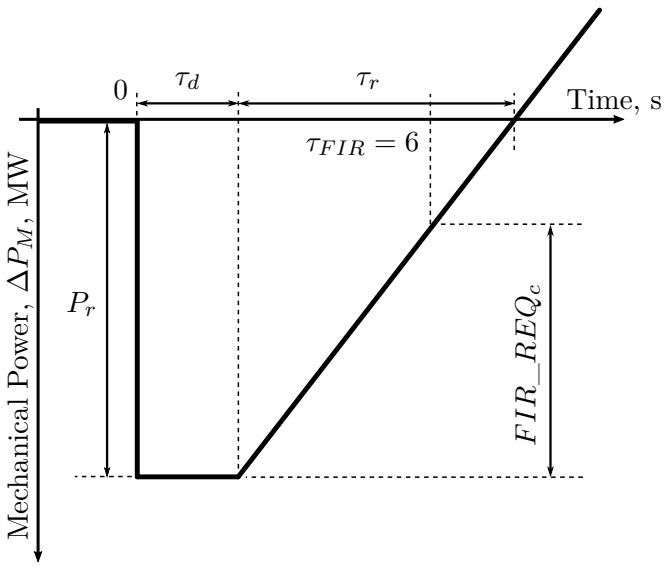


Figure 3.3 The mechanical power transient during a contingency used for creating a model of determining FIR requirements.

A contingency, of size P_r in MW, occurs at time zero. A brief period passes, τ_d in seconds, before reserve starts to respond. The rate at which it responds is defined by τ_r, the time when the risk has to be fully covered by reserve after the delay. Next the swing equation of Eq. 3.7 is modified to include load dynamics with the load damping constant, D:

$$2H\frac{df(t)}{dt} + Df(t) = \Delta P_M(t) \tag{3.9}$$

The response of $\Delta P_E(T)$ is captured in the load damping constant, and by including in the inertial constant, H, a proportion from loads. Solving this differential equation for $f(t)$ when $p(t)$ is set by Eq. 3.8 and the initial condition of 50 Hz at time zero, the frequency transient is obtained. The minimum frequency on the transient is found in Hertz:

$$f_{min} = f_b - \frac{P_r \tau_c f_b}{2H\tau_r}\left(\tau_c \ln\left(\frac{\tau_c}{\tau_r + \tau_c e^{\tau_d/\tau_c}}\right) + \tau_d + \tau_r\right) \tag{3.10}$$

where f_b is the nominal frequency of 50 Hz, and τ_c is the characteristic time constant, $\tau_c = 2H/D$. The method of finding the FIR requirement is by solving for τ_r, when f_{min} is allowed to reach 48 Hz. A closed form equation for τ_r is not possible, so a numerical solver is required, the details of which are not necessary to explain, but it is helpful to simplify Eq. 3.10. It is noticed that the initial RoCoF is found in Eq. 3.10:

$$RoCoF = -\frac{P_r f_b}{2H} \tag{3.11}$$

To further simplify, define τ_m as the amount of time required for the frequency to reach the limit when no reserve responds or load damping occurs:

$$\tau_m = \frac{f_{min} - f_b}{RoCoF} \tag{3.12}$$

Therefore the implicit equation for τ_r is:

$$\tau_m = \frac{\tau_c}{\tau_r}\left(\tau_c \ln\left(\frac{\tau_c}{\tau_r + \tau_c e^{\tau_d/\tau_c}}\right) + \tau_d + \tau_r\right) \tag{3.13}$$

From τ_r the amount of FIR required is how much reserve is provided by 6 seconds, τ_{FIR}, as seen in Figure 3.3 by the parameter FIR_REQ_c, and calculated as:

$$FIR\_REQ_c = P_r\frac{\tau_{FIR} - \tau_d}{\tau_r} \tag{3.14}$$

The impact of inertia and risk on the quantity of FIR required has been obtained through τ_m by Eqs. 3.11 and 3.12. It is of benefit to describe how τ_m relates to τ_r in Eq. 3.13. The unaided time to reach the frequency limit, τ_m, is a good indication of how

much time is available for reserve to respond, $\tau_d + \tau_r$, although not a linear relationship, but rather exponential as seen by the linear line segment in the logarithmic plot of Figure 3.4. This relationship deviates from exponential when τ_m either approaches τ_d or τ_c. The first occurs when the frequency falls quickly; if τ_m approaches the length of delay then it leaves a very short amount of time for reserve to respond, and very large requirement for FIR. The second deviation from exponential occurs when the time to reach 48 Hz and the amount of reserve given by load damping means that it takes a very long time for frequency to reach its minimum. For τ_m greater than τ_c it is not possible to reach the limit. These events characterise normal deviations in power balance, that do not require the same response as needed for contingencies.

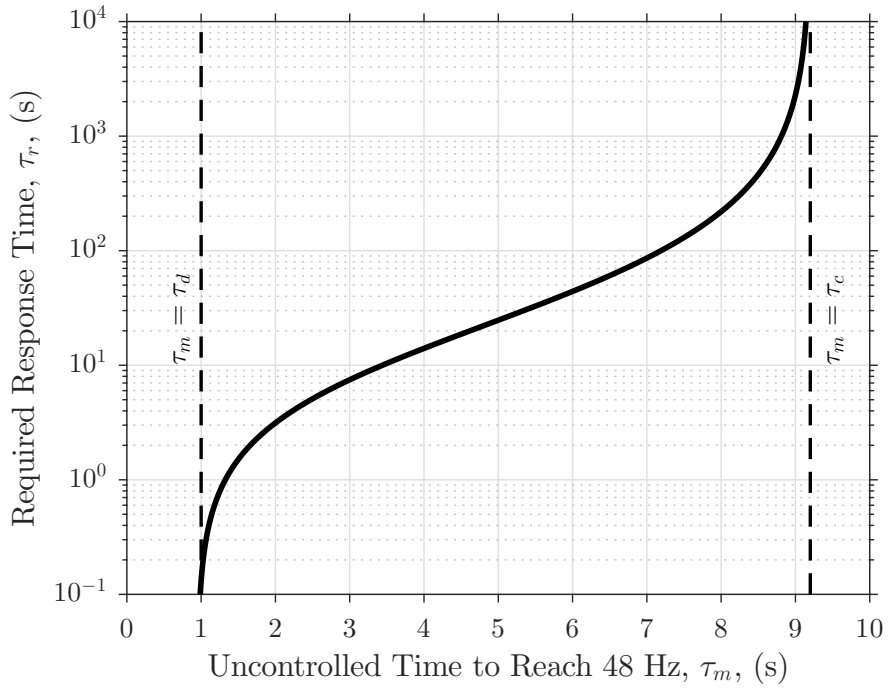


Figure 3.4 The required response time, τ_r, to ensure the frequency does not fall below 48 Hz limit, depending on the uncontrolled time for the frequency to reach 48 Hz, τ_m. The time delay, τ_d, is one second, within the range of capabilities for most hydro power stations. The characteristic time constant, τ_c, is chosen to be 9.2 seconds, it is derived from an inertial constant of 23,000 MW s and a loading damping constant of 100 MW Hz^{-1}. Inertial constant of this size is considered average for the whole of New Zealand, the size of load damping constant is matter of debate, but textbooks recommend a value in this range for a system the size of New Zealand [Kundur et al. 1994].

Since reserve providers require a reasonable amount of time to respond, it is necessary for τ_m to have a lower limit. τ_m should be at least 2 seconds or more, this corresponds to a RoCoF of 1 Hz s^{-1}. For faster events emergency reserve is required, such as AUFLS.

3.4 NEW ZEALAND INTEGRATION STUDIES

The impact of VRE on the demand for IR has been the concern of Transpower and generators, they have investigated these issues, and three reports have been produced that are of interest. These integration studies have analysed the power system for a few scenarios, they consider critical situations were the grid is stressed and have performed dynamic studies. Their findings are briefly reviewed:

- Investigation 5 of the Wind Generation Investigation Project, Effect of wind generation on management of frequency excursions, [Abeyratne 2007].

  This report modelled the loss of Otahuhu B (CCGT) at 340 MW on 25th December 2005 at 4.30 am, the lowest demand trading period over the analysis window, being the worst-case scenario with the least amount of inertia. North Island demand was 1679 MW and 230 MW was being transferred to the South Island across the HVDC link. The power system was modelled using the RMT, which would have been under version 1.0 specification, with a limited HVDC model, not modelling the full benefits of being connected to the South Island and reserve sharing. There were four different scenarios of wind generation, each had a 794 MW wind capacity added, but varied the real-time power output. Each scenario was differentiated by how much wind generation was present at that point in time and by how many Huntly Rankine units were removed (the nominal capacity of 250 MW): scenario 1 was the base case scenario with the historical wind generation of 53 MW, scenario 2 increased wind generation to 189 MW and consequently one Huntly Rankine unit was removed, scenario 3 has two Huntly Rankine units removed (329 MW), and scenario 4 three units (469 MW). The results showed an increased requirement for FIR with the base case starting at just under 200 MW, and for scenario 4 the requirement had increased to just over 250 MW. There were other simulations, such as a higher load scenario for the North Island and a brief consideration of the South Island; however, the results did not show the same increase in FIR required.

- The System Impacts and Costs of Integrating Wind Power in New Zealand, [Strbac et al. 2008].

  This report tries to quantify the indirect costs of wind generation on the power system. These costs were considered to come primarily from three sources: the additional generation capacity required to manage wind generation uncertainty, required transmission capacity, and extra operating reserves. However, the report only considered the capacity adequacy problem and operating reserve requirements. The report anticipated costs for three future scenarios: 313 MW of wind capacity by 2010, 1745 MW by 2020, and 3090 MW by 2030. It estimated that the demand for IR by the 2030 scenario would increase by 184 MW on average

and by 378 MW at maximum, which is a significant increase in demand for IR. However, the analysis makes an inadequate assumption in its estimation of the reserve requirement. It assumes that the demand for reserve is determined by the variability in power output of wind generation in the several minutes time window, and although this is a correct assumption when considering reserve for FK, it is inadequate for IR as the demand for reserves is dominated by the largest contingent risk.

- Technical Advisory Service Contract (TASC) 33, Analysis of Reduced Inertia upon the New Zealand Power System, [SO 2014].

  The report has the greatest accuracy and significance out of the three studies, and is specifically focused on the impacts of reduced inertia upon the RoCoF, and the minimum frequency reached for a contingency. It considers demand for FIR under three snapshots of the grid: winter peak (highest demand), summer peak, and summer trough (lowest demand). For each snapshot, the largest CE was a CCGT operating at 396 MW, 235 MW, and 170 MW respectively. For each snapshot, one thermal or geothermal unit is removed from the base case scenario at a time, and the grid frequency is simulated. The power system was modelled by PowerTech's Transient Security Assessment Tool (TSAT). Utilising the transaction analysis capability tool, the demand for FIR was calculated. The results showed for the winter peak, if 932 MW of thermal generation was removed and replaced by wind generation, then an extra 125 MW of FIR is required. For the summer peak, if 366 MW of thermal generation is replaced then an extra 20 MW is required, and for the summer trough, with 416 MW removed, an extra 51 MW is required. These results show a clear effect of reduced inertia on the requirement for FIR.

Of the three reports, the third is most accurate and relevant, as it is most recent and better encompasses the possible scenarios. Nevertheless after reviewing these methods, it is difficult to determine how accurate taking snapshots of system states reflects all possible situations, and therefore the overall demand for IR. With increased penetration of wind energy, two changes are occurring: inertia is reducing, but risk may be decreasing as well when large thermal units are not required to run. Therefore a method of comparing these two aspects is important.

3.5 MODELLING GENERATION DISPATCH

To estimate how inertia and risk change with different wind energy penetrations, a simplified model of the dispatch process is developed. This model anticipates the likely dispatch of generation at the 30 minute resolution. Starting with a historical dataset of generation information, the model analyses what would happen if 500, 1000, 1500, ... 4000 MW of more wind generation would do to the dispatch. A wind generation time series is obtained from Dougal McQueen's research for the GREEN Grid Project, which uses historic wind speed data, so that correct correlation exists between potential and existing wind farms [McQueen 2016]. This model analyses a historical time period from 2013 to the end of 2015, as this period has the most data available.

The actual generation dispatch is optimised from offers submitted into New Zealand's real-time electricity market to satisfy the electricity demand. Each offer is formed from five blocks of prices and quantities created by the generators. The determination of these price and offer quantity couples is dependent on the nature of the energy resource, the demand for electricity, and the offers from other generators as each participant adjusts their position after each schedule. Therefore these offers are the result of a long planning and scheduling process. Making a single perturbation to the offer set, such as introducing extra wind generation at zero price, would not be accurate as it does not account for changes that would occur to existing offers. The decisions involved in adjusting offers are difficult to emulate, therefore simplifications are made that retain the intentions of participants in the electricity market but avoids a significant amount of complexity by not creating new offer blocks.

In New Zealand, hydro generation is the main source of electricity production. There is only about 4 TW h of storage available in the lakes, which is significantly less than the annual amount produced from hydro generation, which ranges from 22 to 26 TW h. There is only about two months of storage if all inflows were to stop; this creates a significant problem if an extended period occurs between major inflows. Thermal generation is run to manage energy shortage risk. If low lake levels do occur then this is seen as a higher electricity price, as hydro generators increase their prices to avoid being dispatched and more expensive thermal plants replace them.

Thermal generation, from coal and gas in the North Island, has seen a significant decline in the previous decade (2007-2013), as new geothermal and wind generation was built. The expansion of renewable energy stagnated when electricity demand stopped increasing for a while (2007-present). It is anticipated if more wind energy were to be built, then thermal generation will decline again, and the remaining 6 to 10 TW h, depending upon hydro inflows, would further decline as New Zealand satisfies its climate change targets. More information about New Zealand electricity generation and use can be found in Ministry of Business, Innovation, and Employment's (MBIE) electricity statistics.

The dispatch model considers an increase in future wind generation, but it has not made an assumption about future electricity demand for three reasons. (1) In the last decade electricity demand has remained stable as seen in the MBIE statistics. (2) More data is available when using historical information and avoids the difficultly of producing time series. (3) Most importantly, just making changes to wind generation isolates its impacts on reserves, so that arguments about these impacts cannot be attributed to load changes.

The simplified dispatch model makes two key assumptions:

1. For all new wind generation added to the system, it will displace thermal generation.

2. Hydro storage will retain the same lake level and energy shortage risk, but hydro generation can adjust power output to satisfy periods of low wind speed in the short term.

The second assumption is not necessary, because lake storage can be managed by predicting future wind generation, and therefore minimise wind energy curtailment which is likely to occur at higher penetrations. However, guessing the allowable risk of energy shortage that will be acceptable to hydro generators in the future is difficult to calculate. Then minimising deviation from historic lake levels is the easiest method of determining acceptable energy shortage risk. Since future inertia and risk (size of the largest credible contingency) is the key result of this analysis, it is not necessary to focus on minimising wind energy curtailment, and it is left to further GREEN Grid research to solve this problem.

The model starts by separating generation into four types:

1. Must-Run Generation including geothermal, existing wind, and hydro generation with very little to no storage. Whirinaki thermal power station is also included in must-run because it does not run very often.

2. Thermal Generation including the power stations at Huntly, Southdown, Otahuhu, Stratford, Kapuni, and McKee. Otahuhu, Southdown, and one Huntly unit have since left the market, but over the analysis period were still operating. All these thermal power stations are located in the North Island, and the details about each of these power stations can be found in map of North Island in Figure 3.5.

3. Hydro Generation including the schemes of Waikato, Waikaremoana, Waitaki, Clutha, and the single Manapouri power station. The simplified model does not individually separate the power stations within the schemes, as it is assumed generator companies will optimise amongst individual units.

4. Eight scenarios of future wind generation, ranging from 500 to 4000 MW in extra capacity. This is in addition to the 690 MW of existing wind generation included must-run.

Historical data of generated energy in each trading period (30 minutes) is provided by the Electricity Authority, in their Electricity Market Information (EMI) web-page, `emi.ea.govt.nz`. For the eight wind penetration scenarios, energy quantities were derived for each half hour from a power times series of different potential wind farms across New Zealand, these were then combined to achieve the different scenarios.

The first step in the simplified dispatch model is to start with historical dispatch and make changes by replacing thermal generation with new wind generation, so that the total amount of energy produced remains the same to cover the same demand. There are several constraints:

1. Thermal generation is divided into blocks, where a binary decision is made between whether a block remains or is removed. A block exists for each thermal unit, and for each day. A block is comprised of sequence of 48 trading periods starting at 00:00 hours. Separating into blocks recognises that thermal generation cannot repeatedly ramp up or down, nor constantly start up and shut down, as a hydro generator may be able to do. It is not known what cost these operations could incur, so a arbitrary period of day is chosen, providing sufficient duration to avoid repeated cycling, but also allowing for it to be adjusted against wind speed patterns.

2. The quantity of thermal energy removed has to be equal to total new wind energy over an eleven week period. This allows hydro generation to do the final energy balancing for each trading period, while ensuring hydro lake levels do not significantly deviate from historical levels.

3. The choice of which thermal generation block to remove is based on a hierarchy. First it tries to remove all thermal blocks from one unit, then from the next etc. However if it cannot do this because of other constraints it will break this rule, and remove a block from another unit before finishing the first. The hierarchy starts with Huntly Unit 3, Southdown units, Otahuhu, other Huntly Rankine units, McKee, Taranaki combined cycle, Huntly Unit 6, Stratford Peakers, Kapuni, and lastly Huntly Unit 5. The hierarchy starts with units that are mostly likely to be removed from the market, as seen by recent closures. For units that have not since closed a plausible order is made.

4. If there is an over supply of wind energy then wind energy is curtailed in preference to must-run or hydro generation.

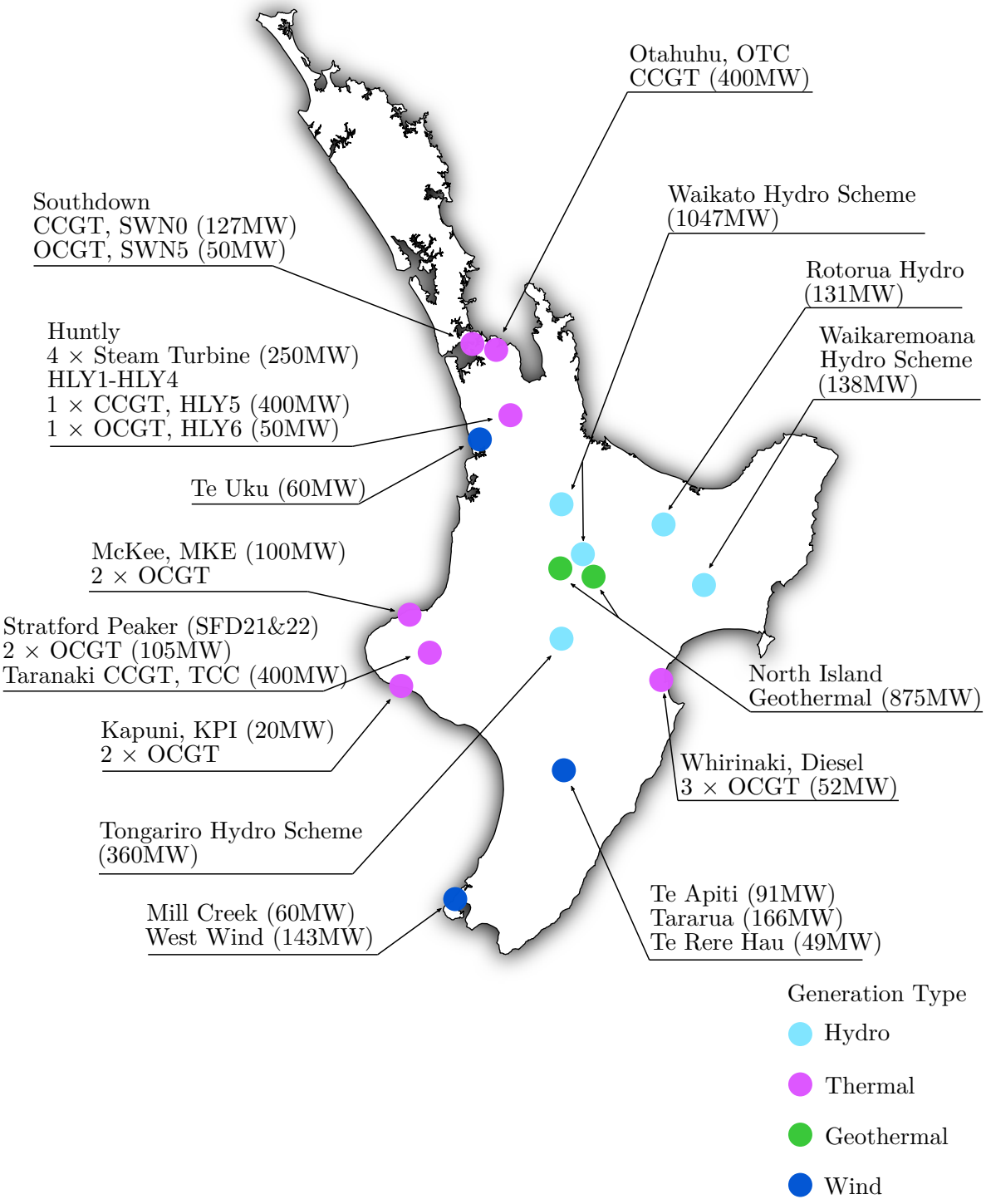


Figure 3.5 Major generators in the North Island, generators and schemes above 50 MW are shown. This image is reproduced and modified from [EA 2018a]. As of June 2019, Southdown and Otahuhu generation have been decommissioned, as well as Huntly Unit 3.

The second step balances energy requirements for each trading period, this is completed by changing hydro generation levels within the limits of available capacity, which are determined by historic offers. The nominal capacity is not chosen, because historic offer limits will better reflect availability. Constraints for the second step are listed as follows:

1. If the change in hydro generation is outside capacity limits, then either thermal or new wind generation are changed to align with hydro generation capabilities.
2. The distribution of changes among hydro schemes is weighted as a function of scheme capacity and energy storage capacity. This ensures a practical change in hydro generation is made, so that schemes with the greater capacity and storage provide the greater effort in balancing.

For further information about the dispatch model, consult the GREEN Grid report [Schipper et al. 2019]. Appendix C of that report provides detail of each balancing step. Appendix D introduces each wind generation scenarios. Appendix E provides more results from the dispatch model.

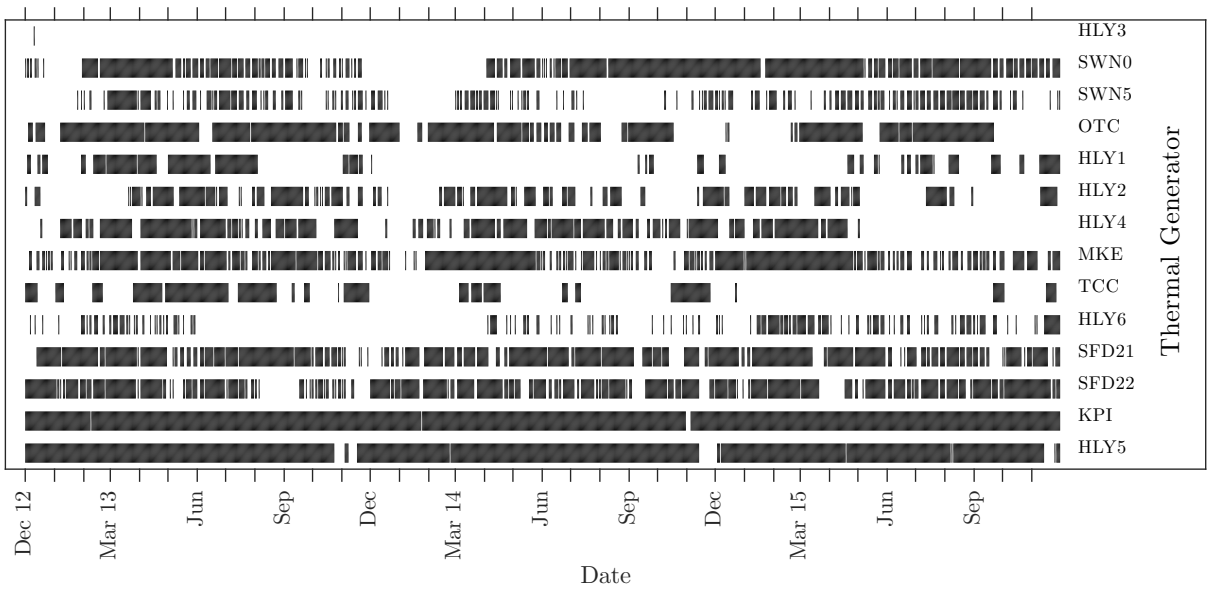


Figure 3.6 Profile of historical thermal generation by day, a black vertical line means that the thermal generator ran for that day, i.e. for at least one trading period in that day it was generating electricity. Day one is for the 1st December 2012, and the last day, 1095 days later, is the 30th November 2015. HLY, Huntly; SWN, Southdown; OTC, Otahuhu; MKE, McKee; TCC, Taranaki Combined Cycle; SFD, Stratford; KPI, Kapuni. More details about each thermal generator can be found in Figure 3.5.

3.5.1 Results from Dispatch Model

The dispatch model produces results that are consistent with expectations. Thermal generation is displaced in proportion to how much wind generation is available, hydro generation does experience wider range of output but is not significantly different from historical operation. More results are presented in the next section on determining inertia and risk, and finally the demand for FIR in the section following.

The historic operation of thermal generators is shown in Figure 3.6. With more wind generation, more thermal generation is removed as shown in Figures 3.7 for 2000 MW and 3.8 for 4000 MW of additional wind generation. There is still demand for thermal generation for particularly dry periods, as seen for 2013 in Figure 3.8 from mid January to the end of March, and for days when there is no wind generation. Even by 4000 MW there are still days when a large number of thermal units are needed, implying that most units are still required even though they provide very little energy. The sporadic cycling of these units, with repeated start ups and shut downs, will also create greater cost for these units, especially to CCGT which prefer baseload operation, as seen in Figure 3.6 for SWN0, OTC, TCC, and HLY5 units, where they tend to run for a large number of days in a row.

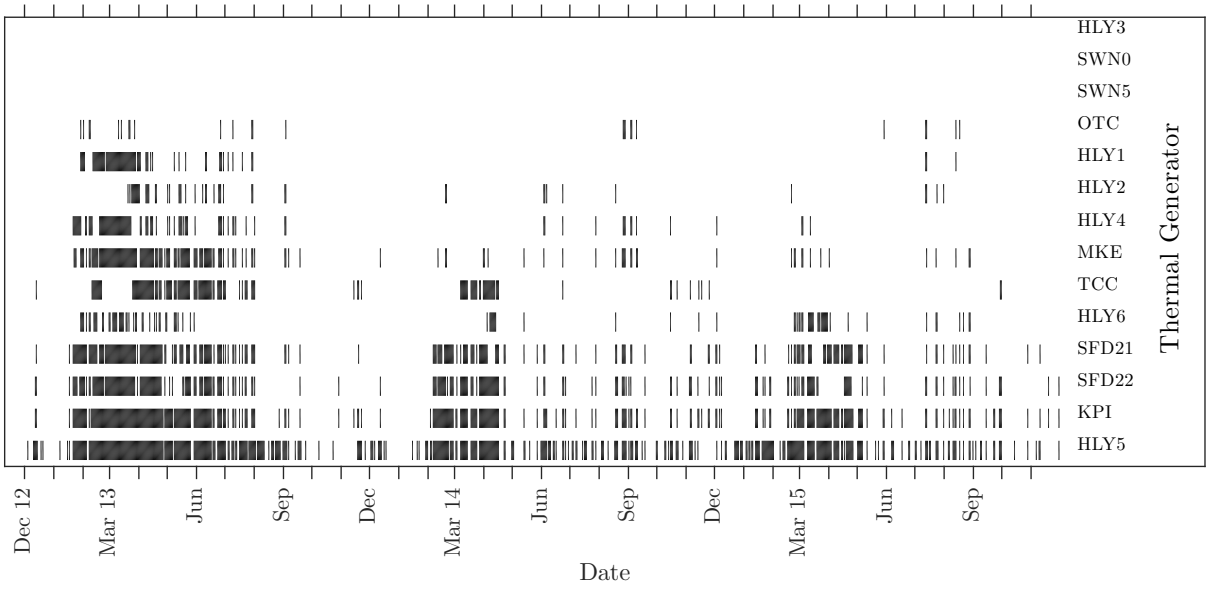


Figure 3.7 Profile of thermal generation with 2000 MW of new wind generation.

The results for hydro generation has seen greater operation around maximum and minimum limits, as expected when managing the variability of wind generation. This is seen in a steeper duration curve, mostly for the Waikato and Waitaki hydro schemes, Figures 3.9 and 3.10 respectively. These duration curves show for what percentage of time the scheme is operating above a certain power output. The limited deviation from the historical curve, means that the hydro schemes are operating in a manner consistent with historic profiles.

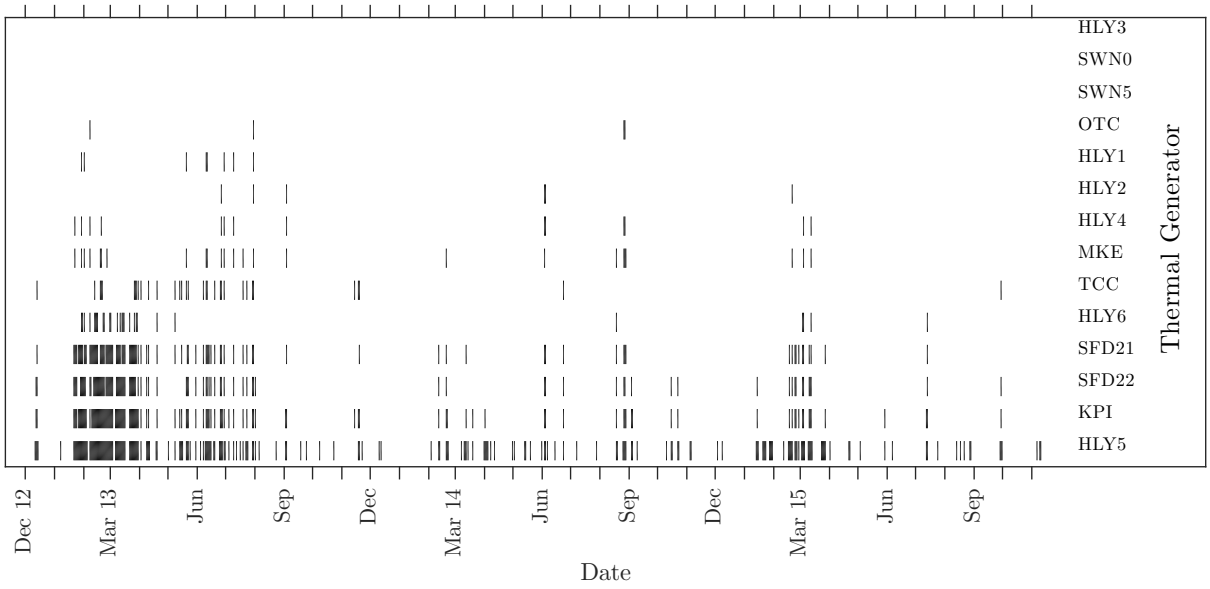


Figure 3.8 Profile of thermal generation with 4000 MW of new wind generation.

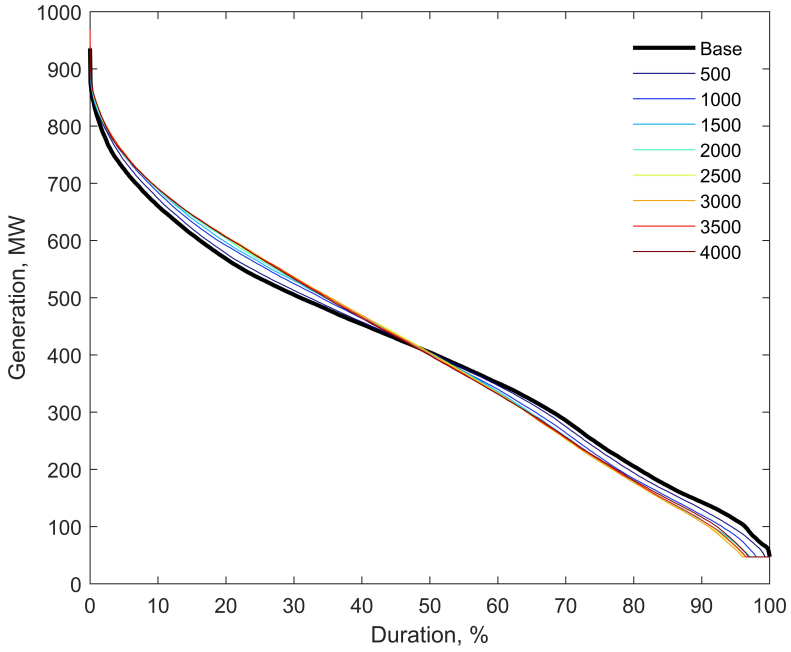


Figure 3.9 Waikato Hydro Scheme Duration Curve

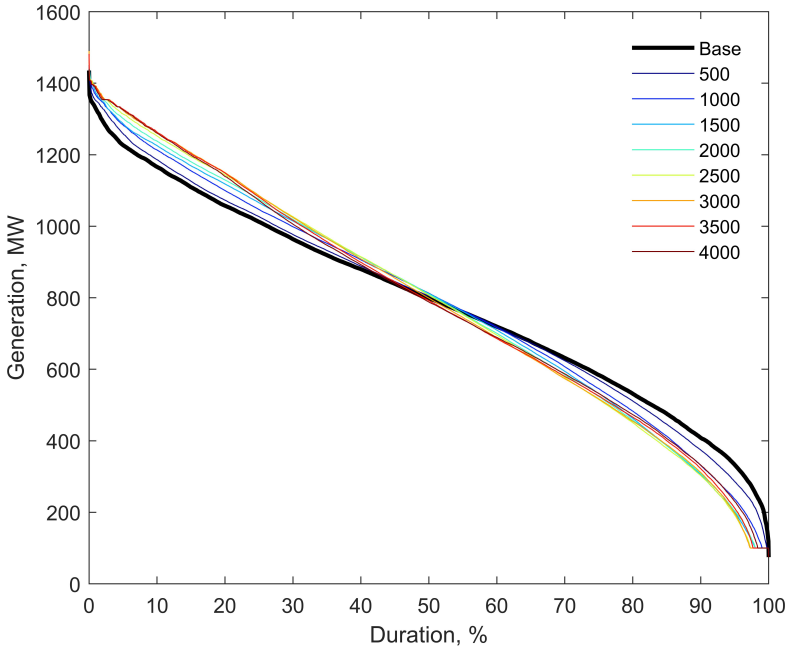


Figure 3.10 Waitaki Hydro Scheme Duration Curve

3.6 IMPACT ON INERTIA AND RISK

From the dispatch model, it is possible to find the total rotational energy from synchronous generation, and the largest credible contingency. This section presents the changes in inertia and risk depending on wind penetration scenario.

The total inertia from synchronous generation is the summation of individual inertial constants of each unit synchronised to the grid. For the historical case, information from Transpower's SCADA system was obtained containing the power output of each unit on the grid. If a unit provided positive power output then that unit was synchronised. The inertial constant of each unit is listed in SO [2014]; some units are missing from this list, so assumptions were made depending on the nominal capacity and generator type.

Reviewing contingent events that occurred during the 2013 to 2015 time period for which the SCADA information was collected, it became apparent that the combined inertial value, obtained by summing individual generator units, under-estimated the actual inertia. It is known that some synchronous condensers were removed from the estimate, and some smaller embedded generators as well, but the main difference is expected to come from motor loads. Therefore making an assumption that the difference is proportional to total electrical demand, the inertial constant was derived for North and South Island loads separately from historical frequency transient information. For the North Island a value of 1.5 MW s/MW was chosen, and the South Island 0.75 MW s/MW. This analysis is presented in Appendix G of the GREEN Grid report

[Schipper et al. 2019]. It is difficult to explain why there is a large discrepancy between each Island, except that a large portion of South Island demand comes from Tiwai Point Aluminium Smelter (5̃40 MW), which requires an AC-DC converter to power the potlines.

The total inertia for new dispatches and wind generation scenarios has to be calculated differently from the historical scenario. Must-run generation remains unchanged in the dispatch model, and so the same inertia time series is utilised. The inertia time series for thermal generation also utilises the same historical time series, but a units contribution for a time period is removed if that unit's block is subtracted from the dispatch. For hydro generation the process is more difficult. The dispatch model only provides a total output for each scheme, which consist of multiple units of varying inertial constant. Decomposing the total output into individual units is not possible as there is insufficient information. Instead a stochastic method is chosen: the likelihood of inertia being a certain value for a given total power output is determined from the historical data. Then a random variable is used to decide the total inertia for a scheme, where the random variable has a distribution in accordance with that likelihood established in the historic data. This process is further explained in Section C.5 of Appendix C in the GREEN Grid report [Schipper et al. 2019].

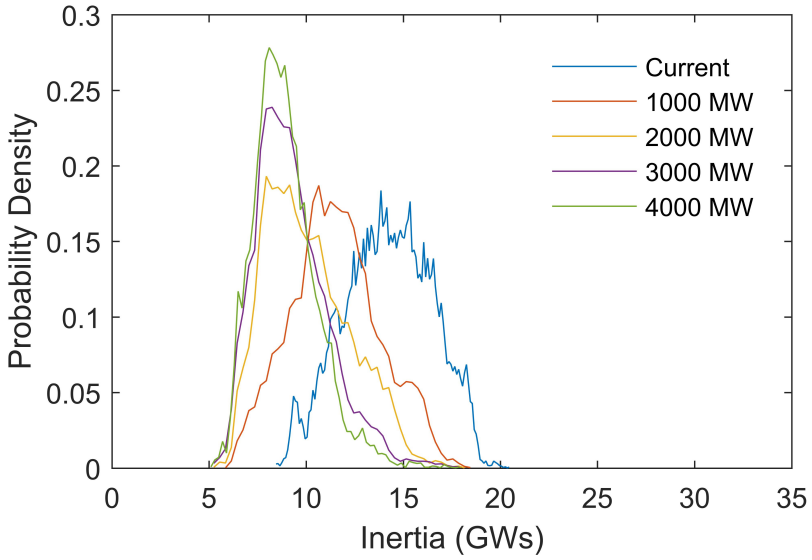


Figure 3.11 The probability density function of North Island inertia, for the historical case (Current), and four of the eight wind generation scenarios. Load inertia is not included.

The results of this analysis are seen in the probability density functions of Figures 3.11 to 3.13, which show the distribution of total inertia from synchronous generators across all operating conditions. These distributions only show the inertia from synchronous generation, the inertia from loads has been omitted. More detailed results are provided in Appendix F of the GREEN Grid report [Schipper et al. 2019], which does

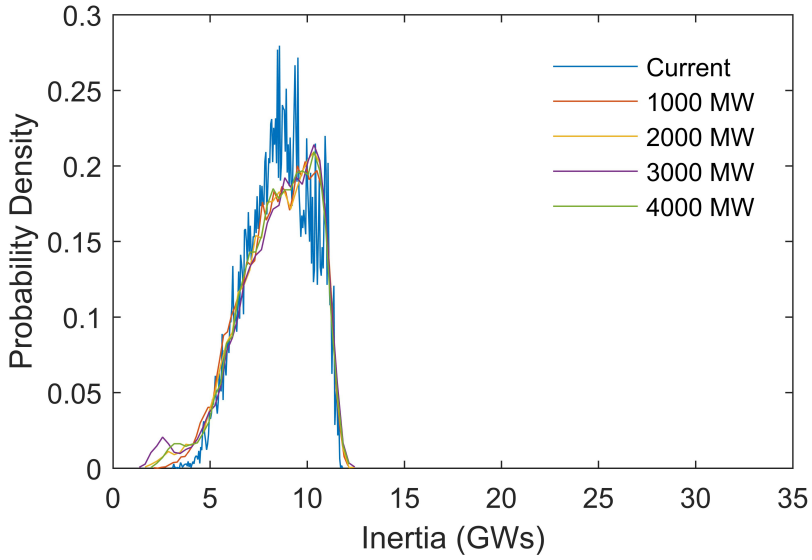


Figure 3.12 The probability density function of South Island inertia, for the historical case (Current), and four of the eight wind generation scenarios. Load inertia is not included.

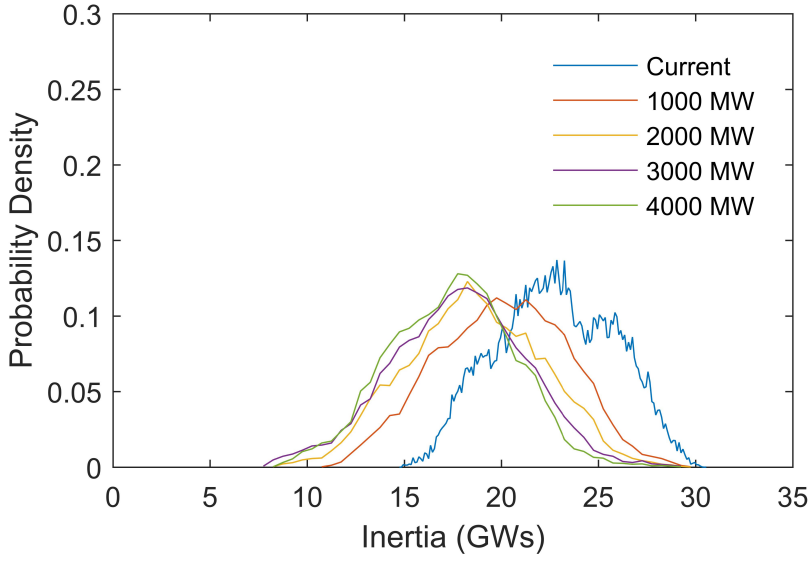


Figure 3.13 The probability density function of total New Zealand inertia, for the historical case (Current), and four of the eight wind generation scenarios. Load inertia is not included.

include the demand side. The distribution of total inertia for the whole of New Zealand is provided in Figure 3.13. Even though North and South Island are not synchronous with each other, the HVDC controls between them do allow for frequency matching and sharing of reserve for contingencies that do not disable the HVDC link. The decline in total inertia is seen mostly in the North Island, Figure 3.11, as thermal generation is replaced by new wind energy. A minimum limit is reached as must-run generation cannot be displaced. The South Island does not see a major change in its total inertia, but with a more spread out profile, the minimum is substantially lower.

Now considering risk, there are two credible contingencies that are analysed: the single largest generator risk, and the loss of a single pole on the HVDC link in north transfer. The largest generator is not distinguished between the two Islands, as is commonly done, because reserve sharing across the HVDC link can simplify the analysis. Although transfer limits do exist for reserve sharing, it is simpler to consider the whole power system as one synchronous system for the loss of a generator. Secondly the largest South Island generator is a 120 MW unit at Manapouri, which is slightly less the 140 MW unit at Nga Awa Purua, a geothermal power station in the North Island. Since Nga Awa Purua is baseload, this unit will usually set the risk over the Manapouri unit, if there is not a larger thermal unit. Thirdly, the HVDC link is usually in operation with both poles for most of the year, except for a standard yearly maintenance of around 36 hours, rarely is reserve sharing not available.

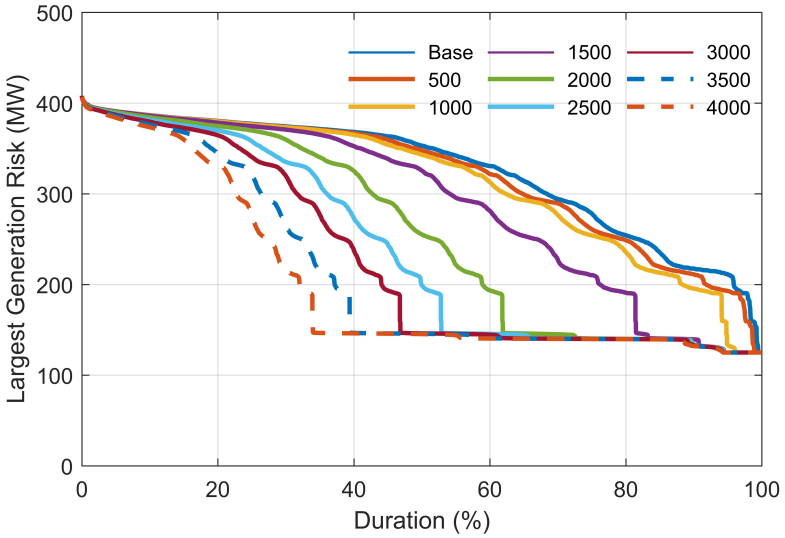


Figure 3.14 The distribution of the largest credible contingency from a generator, represented as a duration curve, and indicating the percentage of time the risk will be above a given value. These results are for the 2013 analysis period.

Determining the largest generator risk is a matter of finding the unit with highest power output. The results from the dispatch model can find this quite easily, as shown in Figure 3.14 by a duration curve. The duration curve remains relatively unchanged

up until 1000 MW of new wind capacity is added. After more wind capacity is added, risk drops significantly as the operation of all CCGTs diminishes.

The HVDC link risk is determined by the transfer of power across the link. If power is transferred north and the connection is lost, then North Island experiences a net reduction in power, and the frequency will quickly drop in the North Island, and quickly rise in the South Island as it experiences a positive power imbalance. The loss of both HVDC poles is considered an ECE, allowing for AUFLS to act, but it is the single pole loss that is of interest as it has to be entirely satisfied by IR.

The total HVDC transfer for each scenario is shown in Figure 3.15. The impact of new wind generation is increased northward transfer of power. Wind generation scenarios have wind farms built in the South Island, but that extra energy is not required there, so it is transferred northwards to replace thermal generation. Clearly the impact is dependent on how many wind farms are built in the South Island, but it is a good assumption that some will be built there to minimise uncertainty of wind generation, recognising the benefits of spatial diversity.

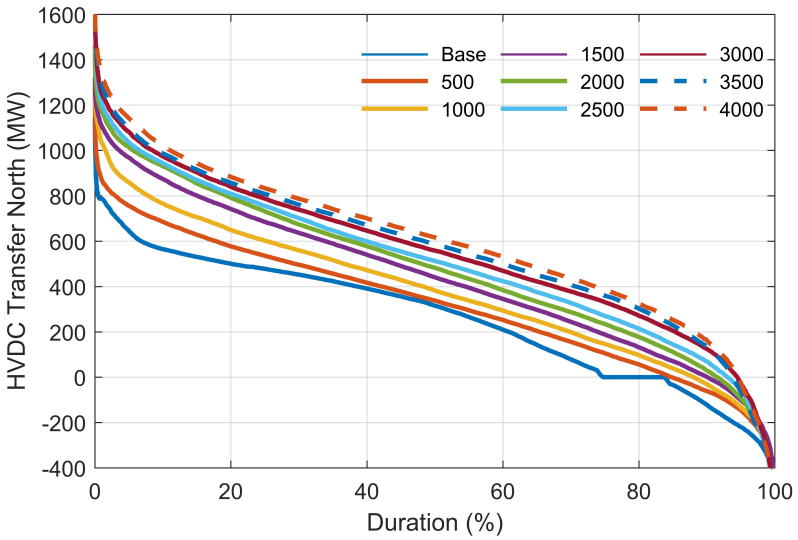


Figure 3.15 Dispatch simulation results for HVDC transfer northwards, a negative value implies a positive southwards flow. The results are for the 2013 analysis period.

It may have been noticed that HVDC transfer has surpassed 1200 MW for periods of time, which is currently the transfer limit. Although the capacity of the converter stations reaches 1400 MW, the cables underneath the Cook Strait limit capacity to 1200 MW. The dispatch model did not limit HVDC transfer, as this was difficult to include in the dispatch model. Therefore in determining the risk, it is assumed that the maximum transfer is 1200 MW.

The HVDC link normally splits power transfer evenly between each pole, but that does not mean the single pole risk is half the total transfer. The HVDC link has the ability to redistribute power from one pole to the other in an emergency. Therefore

the risk of a single pole is the total transfer less 650 MW, as shown in Figure 3.16. As expected the risk increases, and quite significantly compared to historical levels (the curve labelled by 'Base').

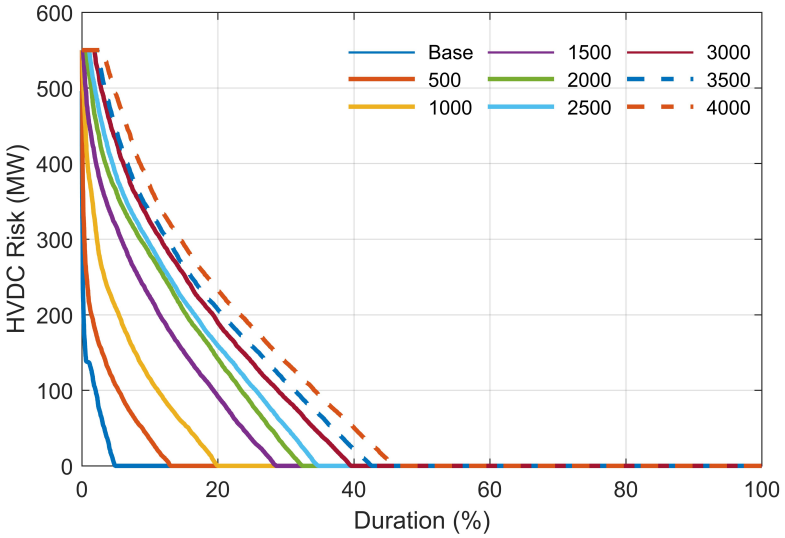


Figure 3.16 Single HVDC pole risk for the different wind generation scenarios. The risk is represented in a duration curve, describing the percentage of time risk above a certain level. The results are for the 2013 analysis period.

3.7 IMPACT ON FIR REQUIREMENT

For these two contingency types, a loss of a single generator and a single HVDC pole loss, the requirement for FIR is determined. Utilising the definitions of Section 3.3, the distribution of τ_m is determined from the times series of inertia and risk. Then Eqs. 3.13 and 3.14 are used to determine FIR_REQ_c, but this requires an assumption about τ_d and τ_c.

The delay time, τ_d, recognises that reserve providers do not respond instantaneously and has an impact on the frequency decline. Each provider has a different delay, so having a single value is an approximation. After reviewing Meridian Energy's (New Zealand's largest generator company) reserve response tests, a one second delay is appropriate for this analysis, this value does not reflect a particular generator of Meridian Energy or an average but is within the range of possibilities. The details of these tests cannot be published, as they are the property of Meridian Energy.

The characteristic time constant of the power system τ_c is dependent on inertial constant, H, and load damping constant, D, so that $\tau_c = 2H/D$. The inertia has been determined already, but an estimate for D is required. Analysing historic contingencies in New Zealand, there was no discernible value for it, implying that $D = 0$. The estimates of this value looked at the initial decline in frequency, trying to estimate the decay in slope before reserve started to respond. Due to insufficient data quality it was difficult to estimate. Kundur et al. [1994], which provides the standard textbook on this subject, comments that this value is between 1 and 2 percent for a general power system: a 2 percent value means for a 1% drop in frequency load will decrease demand by 2%. Therefore with this in mind a value of 0.8% is used.

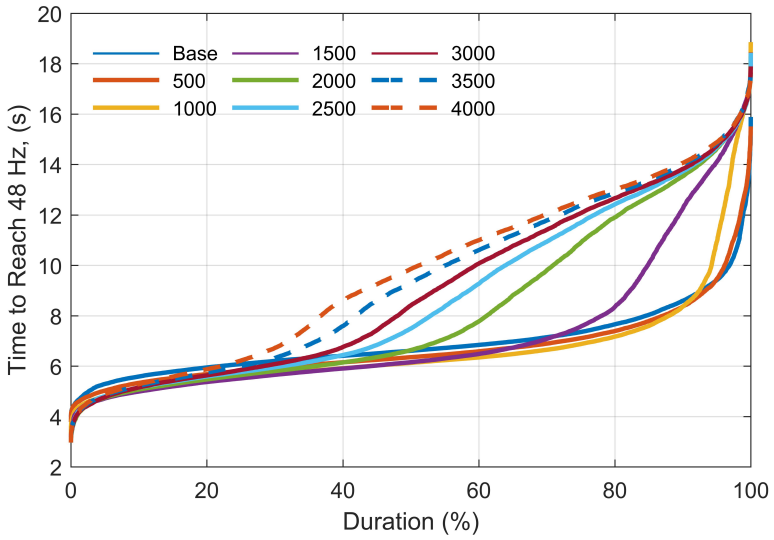


Figure 3.17 Distribution of unaided times to reach 48 Hz, τ_m, for single generator loss contingencies. The results are for the 2013 analysis period.

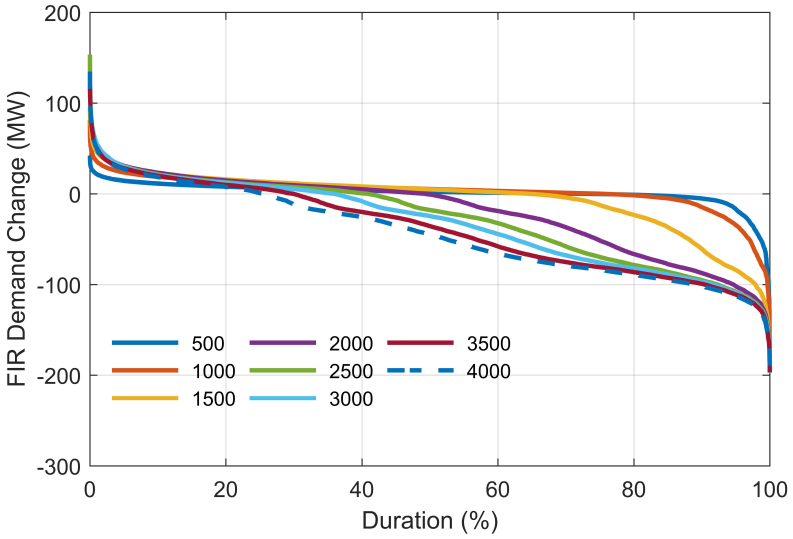


Figure 3.18 Distribution of changes in FIR required from the historical case for single generator loss contingencies. The results are for the 2013 analysis period.

Firstly, a single generator loss is considered. The impact of higher penetrations of wind generation on time to reach 48 Hz, τ_m, without any response to remove the power imbalance is presented in Figure 3.17. The impact on the requirement for FIR is shown in Figure 3.18. Initially the reduced inertia for the first two scenarios (500 and 1000 MW) slightly reduces τ_m from the historical scenario (Base), with a minimal change in requirement for FIR in Figure 3.18. This starts to change for scenarios with 1500 MW or more, as the largest risk starts to decrease in Figure 3.14, the distribution τ_m extends into the higher values to such a point where contingencies are relatively inconsequential, as the demand for FIR reduces. There are brief periods of higher FIR requirements, as seen by small proportions of the curve above the 0 MW in Figure 3.18. The main implication is the greater reduction in reserve requirements.

Now considering a single HVDC pole loss, the distributions of τ_m and changes in the FIR requirement are shown in the duration curves of Figures 3.19 and 3.20 respectively. The results for these events are much more critical. With increasing penetrations, τ_m can fall as low as 2 seconds, and spends much more time at low levels as the penetration level increases, as shown in Figure 3.19 by the widening curves. The requirement for FIR increases significantly, reaching past 1000 MW in some situations in Figure 3.20, an impractical requirement. In New Zealand's real-time electricity market it is likely that HVDC transfer will be reduced to minimise risk in these situations. This will have a cost in wind energy curtailment.

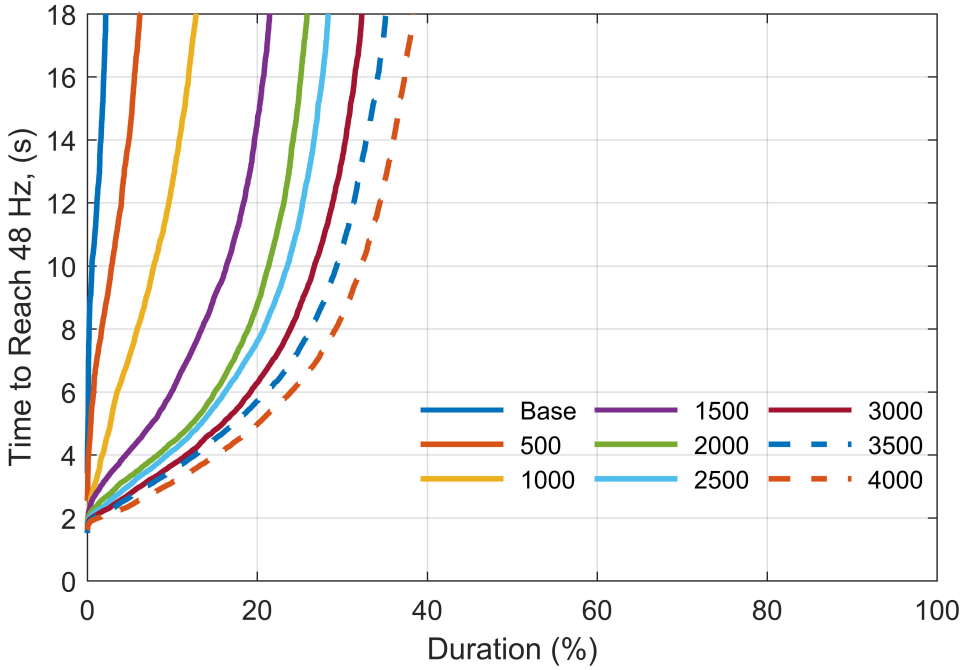


Figure 3.19 Distribution of unaided times to reach 48 Hz, τ_m, for single HVDC pole loss contingencies. The results are for the 2013 analysis period.

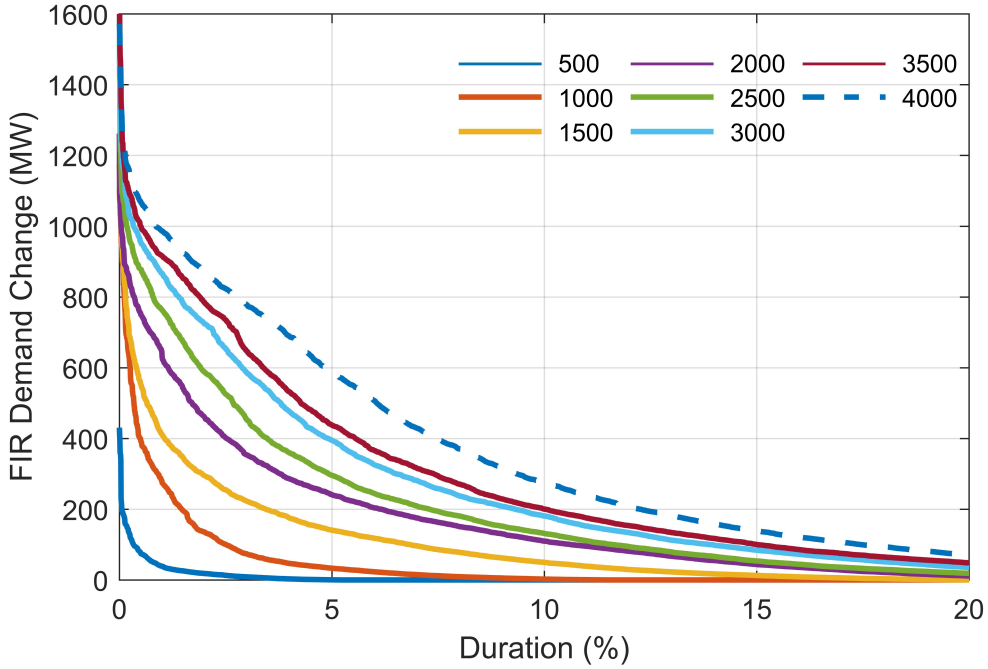


Figure 3.20 Distribution of changes in FIR required from the historical case for single HVDC pole loss. The results are for the 2013 analysis period. The duration scale has been made shorter to remove the longer period of zero demand.

3.8 BALANCING CAPACITY, ENERGY, AND RESERVE COSTS

In the future, under higher penetrations of wind generation, the risk of a single HVDC pole loss is going to be the most important factor that defines requirements for IR and capacity requirements in the North Island. If it is desired that thermal generation be removed in the North Island and replaced by wind energy, then another source of definite capacity is required, as wind cannot be relied upon to be blowing all the time. The South Island is an excellent source of capacity, especially its hydro generation, and to a lesser extent potential wind generation, as there is a likelihood that the wind will be blowing in the south when it is not in the north. However if every MW of power transferred north across the HVDC link also requires another MW of reserve in the North Island, the benefit of South Island capacity is lost.

There is a clear difficulty in achieving 100% renewable power system in New Zealand, the option of just replacing thermal generation with wind generation does not appear to be cost effective. Each new MW of wind capacity becomes less effective, as the chances of that unit of energy being curtailed rises. This effect is minimised by spreading out wind turbines across a larger geographical area, but is hampered by large reserve requirements to cover HVDC risk. Therefore other options need to be considered to minimise thermal generation:

- Build more power stations from more certain sources of renewable energy like hydro and geothermal in the North Island.
- Build more reserve capacity and energy storage capacity in the North Island from energy storage technologies such as pumped hydro, batteries, and hydrogen.
- Build a second HVDC link between the North and South Islands. This increases access from South Island generation to the North Island, while minimising risk.

It may have to be accepted that some thermal generation is going to be needed to minimise costs. This requires a robust economic analysis of all options. The size of this problem is beyond the time and space of this thesis. This thesis focuses on developing a tool that can quickly compare IR providers and accurately predict IR requirement, so that the optimal decision for the dispatch of reserve is made. This tool is developed so that it can be integrated into a market optimisation to find the best dispatch of resources. This tool would then be able to help analyse the cost of different options of achieving 100% renewable electricity in New Zealand, and potentially other power systems as well.

3.9 DIFFICULTIES WITH CURRENT RESERVES OPTIMISATION

The demand for IR is expected to change, as shown in Section 3.7. The average requirement is likely to decrease, but the peak requirement is anticipated to be higher. This increases the demand for reserve capacity, but reduces the time it is utilised. Coupled with decreasing inertia, the range of possible conditions the IR market experiences is wider. Knowing that the current market optimisation is formulated as a MILP, the linearised IR constraints will suffer from several drawbacks as follows:

1. The definition of FIR at 6 seconds for SR results in over procuring reserve in low inertia and high risk situations. This is to ensure sufficient ramp rate so that the frequency remains above 48 Hz, but in doing so it removes capacity from the market that could provide energy, which may be critical in highly constrained periods.

2. An assumption is made in the optimisation formulation that one MW of risk is equivalent to one MW of IL or SR. Therefore the decision to increase both risk and reserve, or decrease both, is by comparing the marginal cost of risk with the marginal cost of reserve directly. However due to the finite time it takes reserve to respond, the impact of increasing risk by one unit may be having to increase reserve by more than one MW. Therefore the inherent preference should be first to minimise risk, increase IL, and then the extra capacity made available from SR can be utilised for energy output, which will increase the rotational energy on the power system. Therefore one to one comparison is suboptimal.

3. An equal comparison between IL and SR, even though they are defined differently and have different dynamic performance, may not provide the right incentive for investment. This may over-value SR and under-value IL which may result in inefficient outcomes. Secondly, knowing that average reserve requirement will decrease and the peak reserve requirement will increase, ensuring sufficient peaking reserve will be critical. If the right investment signals are not there then this may not be possible.

These concerns are not original to this research. In New Zealand, Transpower understands the impact of low inertia on FIR requirement [Pelletier et al. 2012]. ENERNOC, a large provider of IL in New Zealand is concerned with low inertia, and the need for appropriate reserve resources, recommending a separate market for IL [ENERNOC 2015]. Transpower has analysed the benefits of potential upgrades and new products to the IR market [Phethean et al. 2015a].

The ideal tool is one where the performance of each provider is captured in the dispatch optimisation, not having to define reserve as the output at one particular

time, but with more generality. This should make it possible to uniquely value reserve products. The portfolio of products should be optimised to stop the frequency dropping below its limits, thereby avoiding the over procurement of reserve. The goal of this thesis is to develop this tool.

Chapter 4

LITERATURE REVIEW

Chapter 3 has shown the motivation for refining definitions of IR in New Zealand, particularly to avoid over procuring FIR to ensure sufficient speed of response. New definitions should have reference to performance characteristics, and be constrained against frequency limits, fulfilling the purpose of reserve to arrest frequency decline in emergencies. Other researchers have noticed this problem and have developed solutions, this chapter is a review of these methodologies. The literature is separated into two groups: constraints consistent with MILP formulations, and Optimal Control, which is a broad area of research and is only briefly reflected on.

4.1 MILP FORMULATIONS

Motivated by increasing penetrations of VRE, the literature has considered three types of constraint in forming eletricity markets: limit on RoCoF, minimum frequency, and minimum steady state frequency. The difficulty in incorporating these constraints is linearising them, so that they are consistent with MILP. The minimum frequency constraint is of particular interest as this is derived from differential equations. This section describes optimisation formulations in the literature and compares key characteristics. The main focus of this review is on reserves and their constraints, but any consideration of reserves also entails considering energy optimisations as well, even if they are not co-optimised together. Therefore it is assumed that the energy optimisation is consistent with MILP requirements, such as from current electricity markets, and the reserve constraints are analysed without reference to a particular electricity market. The three new types of reserve constraints are analysed in isolation, even constraints that determine the sharing of generation capacity between energy and reserves are omitted from the discussion, as well as the objective function. Therefore, no energy constraints, i.e. related to the energy market, are shown in this section.

4.1.1 Rate of Change of Frequency Constraint

RoCoF constraints are proposed to avoid violation of equipment capabilities. New Zealand does not have a standard on equipment RoCoF capability, but a standard has been developed in Ireland limiting it to -0.5 $\mathrm{Hz\,s^{-1}}$ [EirGrid 2015], which is being extended to -1.0 $\mathrm{Hz\,s^{-1}}$ [CRU 2018]. Frequency has its fastest descent just after a contingency has occurred, before any reserve response starts to reduce the power imbalance. Therefore to reduce the maximum deviation in RoCoF, it is important to minimise the initial descent, which is dependent on the initial loss of generation and inertia. RoCoF is defined by the following equation:

$$RoCoF = \frac{df}{dt} = \frac{f_b \Delta P}{2HS_b} \tag{4.1}$$

where f is the frequency in Hz, f_b is the base frequency and is equal to the nominal value of 50 Hz in New Zealand. ΔP is the imbalance between the combined mechanical power supplied and the electrical power being drawn from the synchronous machines. S_b is the base power, and defines the relationship between the inertia, H, in seconds and the inertia in conventional units, J, in $\mathrm{kg\,m^2}$:

$$H = \frac{J\omega_b^2}{2S_b} \tag{4.2}$$

where ω_b is the nominal angular velocity of the rotor, $\mathrm{rad\,s^{-1}}$. For simplicity a two pole generator is assumed, and $\omega_b = 2\pi f_b$. Eq. 4.2 provides a helpful interpretation of H as the rotational energy stored in the rotating parts divided by the base power. Therefore HS_b is equivalent to the amount of kinetic energy stored in all synchronous generator and associated turbine. Eq. 4.1 assumes that the initial frequency is the nominal frequency of 50 Hz, however if the initial frequency is slightly different, then Eq. 4.1 is adjusted to:

$$RoCoF = \frac{df}{dt} = \frac{f_b \Delta P}{2HS_b} \frac{f_b}{f_0} \tag{4.3}$$

where f_0 is the initial frequency. When the initial frequency is lower than the nominal frequency, it has the effect of increasing the RoCoF slightly more than what would be expected at nominal. This is due to the squared relationship between energy and angular velocity, $E \propto \omega^2$. In New Zealand the initial frequency before an event can be anywhere between 49.9 and 50.1 Hz. Using the lower value, the initial frequency of 49.9 Hz has the effect of increasing the RoCoF by 0.2%. The initial frequency has minimal impact on the RoCoF and so Eq. 4.1 is sufficient, as the inaccuracy in predicting the contingency risk, ΔP, and the inertia, H is greater than the potential impact of the initial frequency.

Eq. 4.1 is linearised by noting that the RoCoF constraint is static, and is rearranged

as follows:

$$HS_b\Big(\frac{2RoCoF_{min}}{f_b}\Big) \leq \Delta P \tag{4.4}$$

This equation is linear as it is only the inertia, H, and the size of the contingency, ΔP, that are variables. $RoCoF_{min}$ and ΔP are negative values. Sometimes it is common to simplify HS_b to kinetic energy, KE, and referring the parameter to the right hand side [Doherty et al. 2005]:

$$KE \geq \Big(\frac{f_b}{2RoCoF_{min}}\Big)\Delta P \tag{4.5}$$

KE is the total kinetic energy from all synchronous generators after the contingent event. Considering only N-1 contingencies where a single generator is lost, then for a set of potential risks called GenRisk, then there is a series of RoCoF constraints as shown in [Doherty et al. 2005]:

$$\sum_{i \neq k} KE_i \geq \Big(\frac{f_b}{2RoCoF_{min}}\Big)P_k - KE_L \quad \forall k \in GenRisk \tag{4.6}$$

where KE_i is the kinetic energy of each individual generator, KE_L is the kinetic energy of loads, and P_k is the pre-contingent output of the potential risk.

4.1.2 Minimum Frequency Constraint

1. *Doherty, 2005*

   Doherty et al. [2005] provides the first approach to linearising the minimum frequency constraint. This is done in the context of the Irish power system. The frequency response of the Irish system can be modelled in great detail, but for computational restrictions it has been simplified to the form shown in Figure 4.1. The parameter of the reserve response, α, is tuned against results from a detailed model simulation for events that result in a minimum frequency of 49.3 Hz, the minimum frequency for contingencies in Ireland. Results derive a formula for α based on the amount of reserve procured after five seconds, R:

   $$\alpha = 2.869R - 8.64 \times 10^{-4}R^2 \tag{4.7}$$

   For reserve requirements ranging from 200 to 400 MW, it is the proportional term of Eq. 4.7 that has the greater impact on the reserve response gain. For example, if 403 MW of reserve is procured, α is 1020 MW Hz^{-1}, then the steady state frequency should reach 49.6 Hz. It is due to the finite response time that the frequency first drops to 49.3 Hz in Figure 5 of [Doherty et al. 2005].

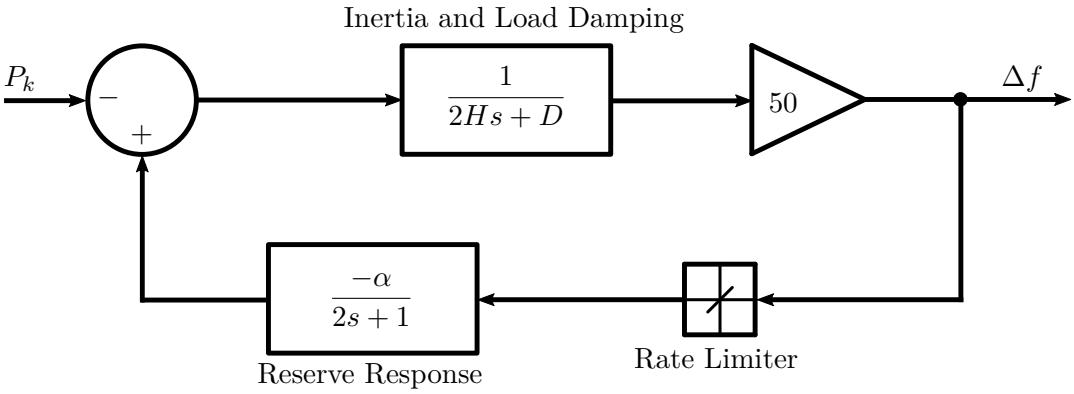


Figure 4.1 Model of Ireland's frequency response obtained from Doherty et al. [2005]. The units of the input power, P_k, is in MW and models the loss of generation during a contingency. The output frequency, Δf, is the perturbation in frequency from the nominal 50 Hz. The gain of 50 after the inertia and load damping model converts the frequency from a value in per unit to one in Hz. The input and output variable units necessarily imply that the units of inertia, H, are in MW s, the load damping constant is in MW/pu frequency. The reserve response gain, α, is in MW Hz^{-1}. The rate limiter is set to 0.12 $\mathrm{Hz\,s}^{-1}$. The details about how values for inertia and load damping are obtained can be found in [Doherty et al. 2005].

The minimum frequency constraint for Figure 4.1 required simulating different scenarios. Multiple simulations are needed to cover the different conditions of the Irish system. In total, over 20,000 simulations determine how inertia, electrical demand, contingency size, and reserve set the minimum frequency. Since the model is only valid for minimum frequencies that approach close to 49.3 Hz, only results from these simulations are kept, then a set of linear inequalities is fitted to it. Therefore the condition on inertia, demand, contingency size, and reserve to keep the minimum frequency above 49.3 Hz in the optimisation are:

$$\sum_{i \neq k} KE_i \geq C_{j,1}L + C_{j,2}P_k + C_{j,3}\sum_{i \neq k} R_i + C_{j,4} \quad \forall k \in GenRisk \quad \text{and} \quad \forall j \tag{4.8}$$

where C is a matrix with coefficients:

$$C = \begin{bmatrix} -5.33 & 230.93 & -124.89 & -834.79 \\ -5.56 & 248.35 & -150.80 & -580.40 \\ -4.69 & 182.40 & -67.61 & -832.85 \\ -4.84 & 198.17 & -76.98 & -2887.29 \\ -4.48 & 175.78 & -52.08 & -4143.23 \end{bmatrix} \tag{4.9}$$

These equations are taken directly from [Doherty et al. 2005]. L is the electricity demand and factors into the inertia and load damping estimations. Little detail is provided on how these linear constraints were produced other than optimal curve fitting was completed in MATLAB.

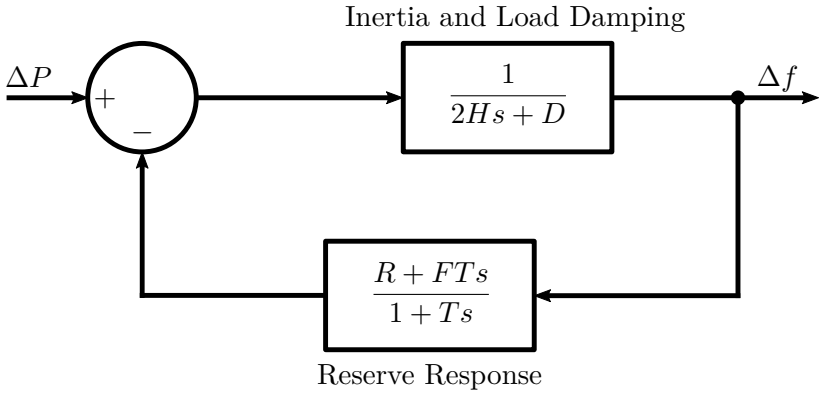


Figure 4.2 Model of the frequency response of a power system used by Ahmadi and Ghasemi [2014]. Frequency and power quantities are both in pu. The base frequency and base power are left undefined.

2. *Ahmadi, 2014*

Ahmadi and Ghasemi [2014] set out to linearise an analytical model of the minimum frequency for a contingency, this is possible by solving a set of linear differential equations and finding the minimum in the resultant frequency transient. The model is shown in Figure 4.2.

For a step response, due to a loss of generation, ΔP, the frequency deviation is defined by the following equation:

$$\Delta f(t) = -\frac{\Delta P}{2HT\omega_n^2}\left(1 + \frac{e^{-\xi\omega_n t}}{\omega_r}\left(\omega_n^2 T \sin(\omega_r t) - \omega_n \sin(\omega_r t + \phi)\right)\right) \tag{4.10}$$

where

$$\omega_n = \sqrt{\frac{D+R}{2HT}} \tag{4.11}$$

$$\xi = \frac{1}{2}\frac{2H + T(D+F)}{\sqrt{2HT(D+R)}} \tag{4.12}$$

$$\omega_r = \omega_n\sqrt{1-\xi^2} \tag{4.13}$$

$$\phi = \sin^{-1}(\sqrt{1-\xi^2}) \tag{4.14}$$

The time when the minimum frequency occurs is solved by taking the derivative of $\Delta f(t)$ and find the time it equals zero:

$$t_{min} = \frac{1}{\omega_r}\tan^{-1}\left(\frac{\omega_r}{\xi\omega_n - 1/T}\right) \tag{4.15}$$

Substituting the minimum time, Eq. 4.15, into Eq. 4.10 the minimum frequency is obtained.

$$\Delta f_{min} = -\frac{\Delta P}{R+D}\left(1 + e^{-\xi\omega_n t_{min}\sqrt{\frac{T(R-F)}{M}}}\right) \tag{4.16}$$

In [Ahmadi and Ghasemi 2014], the minimum frequency constraint requires that $\Delta f_{min} \geq \Delta f_{limit}$, adding this to Eq. 4.16 and rearranging:

$$-(R+D)\frac{\Delta f_{limit}}{\Delta P} - e^{-\xi\omega_n t_{min}}\sqrt{\frac{T(R-F)}{M}} \geq 1 \tag{4.17}$$

To formulate this constraint into a MILP optimisation problem, it is necessary to formulate Eq. 4.17 as a series of linear constraints. Ahmadi and Ghasemi [2014] provides a methodology to fit a series of hyper planes, but for this to be easily achieved, it is required that the left hand side of Eq. 4.17 be a convex function with R, F, and M variables. However this has not been shown. It may even be possible to include ΔP as a variable in order for the optimisation to optimally choose when to reduce the largest credible contingency, but [Ahmadi and Ghasemi 2014] has not done this. Guggilam et al. [2018] has also derived a minimum frequency for a similar second order model, but with different parameters. However, there is a fundamental difference in optimisation: [Ahmadi and Ghasemi 2014] is incorporating frequency dynamics into the dispatch process, whereas [Guggilam et al. 2018] is optimising inertia and droop parameters of Distributed Energy Resources (DER) after the dispatch.

3. *Chávez, 2014*

Instead of a linear system model, Chávez et al. [2014] considers the minimum frequency constraint by limiting dynamics to inertia, a single constant ramp rate for reserves, and the deadband of generator governors. The analysis starts by considering the minimum required ramp rate for the minimum frequency to be above the limit. This limit is set by the trigger frequency for emergency load shedding. Dynamics are determined by the swing equation:

$$M\frac{df(t)}{dt} = \Delta P(t) \tag{4.18}$$

where M is the mechanical time constant and is equal to $2H$ from Eq. 4.1, units are in MW s. f is the frequency in per unit, the nominal frequency is 60 Hz in accordance with the ERCOT power system. ΔP is the power imbalance between supplied mechanical power and electrical load in MW. The power imbalance for a contingency is shown in Figure 4.3.

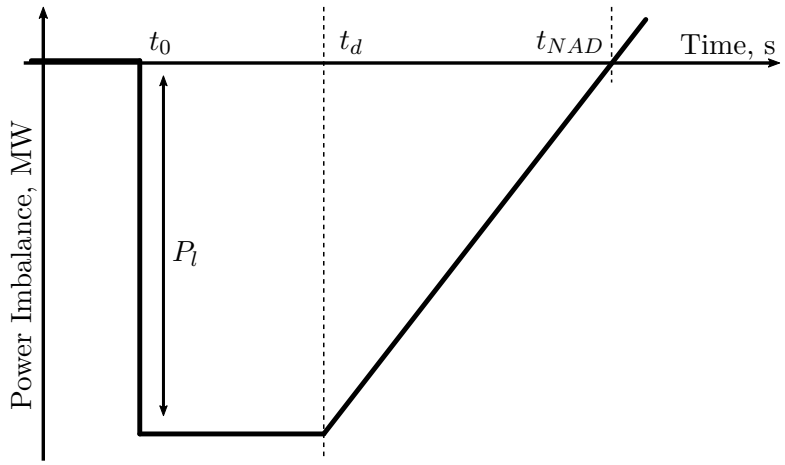


Figure 4.3 The power imbalance as result of loss of generation, P_l. After a short time delay, governor action responds by increasing power back to pre-fault conditions and arresting the frequency drop at t_{NAD}. The time delay, t_d, is caused by the governor deadband. Reserve responds with a combined ramp rate of C_{NAD} in $\mathrm{MW\,s^{-1}}$.

By integrating Eq. 4.18 starting at the time of the contingency, t_0, to the time at which the frequency reaches its minimum when $\Delta P = 0$ again, t_{NAD}, the minimum frequency is obtained:

$$f_{min} - f_0 = -\frac{1}{M}\Big(P_l t_d + \frac{P_l^2}{2C_{NAD}}\Big) \tag{4.19}$$

This is the same as Eq. 5 from [Chávez et al. 2014], where f_0 is the initial frequency before the contingency and f_{min} is the minimum frequency, both these values are in per unit and measured in deviation from unity. To incorporate the governor deadband, the constant Rate of Change of Frequency before generators start responding is used to predict when the frequency is first outside the governor deadband:

$$t_d = \frac{M}{P_l} f_{db} \tag{4.20}$$

where f_{db} is a positive value for the deadband frequency in per unit. Therefore substituting Eq. 4.20 into Eq. 4.19 and rearranging for C_{NAD}, the minimum requirement for ramp rate is obtained:

$$C_{NAD} \geq \frac{P_l^2}{2M(f_0 - f_{min} - f_{db})} \tag{4.21}$$

The minimum ramp rate requirement, C_{NAD}, also puts a constraint on the maximum amount of time that reserve should be allowed to respond to an event, that is all reserve used to stop the decline in frequency should respond before $t_{resp} = P_l/C_{NAD} = t_{NAD} - t_d$ after it starts responding. To formulate these

constraints into a unit commitment problem, two inequality equations are required to ensure the minimum frequency does not go below the limit:

$$\sum_i R_i \geq P_l \tag{4.22}$$

$$R_i \leq c_i t_{resp} \tag{4.23}$$

where R_i is the amount of reserve procured from unit i in MW, and c_i defines the maximum ramp rate that a unit can sustain, in $\mathrm{MW\,s^{-1}}$. Eq. 4.22 ensures that enough reserve is procured that the frequency minimum is reached and frequency starts increasing. Eq. 4.23 ensures that all the procured reserve has had sufficient time to reach its dispatched amount before the frequency is expected to reach its limit. The result is faster resources are dispatched first when inertia is reduced and t_{resp} is shorter.

4. *Wen, 2016*

   The use of t_{resp} in Eq. 4.23 as a parameter determined by the system conditions provides a linear constraint that can be easily solved in MILP formulations. However this formulation suffers from one important limitation, it does not include the maximum contingency size, P_l, as a variable in the optimisation. Because if it was included, after substituting $t_{resp} = P_l/C_{NAD}$ into Eq. 4.21 by removing C_{NAD} and rearranging, it is found that $t_{resp}P_l \leq const$ which is non-linear. Therefore [Wen et al. 2016] proposes a technique of linearisation to remove this limitation and still solve the problem as MILP. The linearisation process is the Reformulation-Linearization Technique. Wen et al. [2016] also includes the capabilities of Battery Energy Storage systems into the optimisation.

5. *Sokoler, 2016*

   Previous work has tried to simplify the modeling of reserve provision and incorporate the minimum frequency constraint into MILP through a linearisation technique. The differences between these approaches have been whether reserve responses should be modelled through a linear differential equation or by a constant rampimg, which is constrained by a maximum ramp rate. The approach of Sokoler et al. [2016] is to retain detailed models of how reserve is provided, decouple the relationship between grid frequency and reserve response, and have cautious constraints to ensure that the minimum frequency during a contingency is greater than the limit. This approach is different because it retains the accuracy of how reserve is provided, whereas other models have simplified, but loses this accuracy elsewhere in being over conservative in ensuring its minimum frequency constraint is satisfied.

The approach starts by considering that the minimum frequency constraint can be reformulated as a minimum rotational energy constraint, as frequency and energy are related, $E = (J\omega^2)/2$, where rotational energy, E, is proportional to the angular velocity squared, ω, which for synchronous machines is directly coupled to the grid frequency. Therefore to keep $f(t) \geq f_{limit}$ is to ensure that the power imbalance as a result of a contingency does not cause a change in rotational energy to go below the limit:

$$\Delta E_{lim} = HS_b\left(\frac{f_{limit}^2 - f_{nom}^2}{f_{nom}^2}\right) \tag{4.24}$$

f_{nom} is the nominal frequency in Hz. H and S_b retain the same definitions from Eq. 4.2 and multiplied together is the stored rotational energy at nominal frequency. Since the frequency limit is less than the nominal then the energy limit is negative. The energy provided by reserve, $E_R(t)$, less the energy lost due to the loss of power, P_l, needs to be greater than the energy limit:

$$E_R(t) - P_l t \geq \Delta E_{lim} \tag{4.25}$$

Total energy from reserve is summed from individual providers. Energy is integrated from the power output, $R_i(t)$, for each unit i:

$$E_R(t) = \int_0^t \sum_i R_i(\tau) d\tau \tag{4.26}$$

It is recognized by [Sokoler et al. 2016] that the relationship between grid frequency and reserve is coupled: frequency is dependent on the power balance between the reserve and lost power, the frequency in turn determines the response from providers. This keeps the power system stable, but is difficult to model. Sokoler et al. [2016] generalised the coupling by a series of implicit differential equations:

$$\mathcal{F}(R_T, f, \dot{f}) = \dot{f}(t) - \frac{f_{nom}^2}{2HS_b f(t)}(R_T(t) - P_l) = 0 \tag{4.27}$$

$$R_T(t) = \sum_i R_i(t) \tag{4.28}$$

$$\mathcal{G}_i(f, R_i, \dot{R}_i, \cdots, (R_i)^n) = 0 \tag{4.29}$$

Eq. 4.27 is the non-linear version of the swing equation. Eq. 4.29 describes the reserve response of each provider, but does not specify the control structure as the model is dependent on several non-linear processes. The implementation of Eq. 4.27 and 4.29 is difficult to incorporate into MILP. Therefore the approach is to

solve Eq. 4.29 for a predetermined frequency transient, a linear decreasing function from f_{nom} to f_{limit} in time t_c is used. If the choice of reserve options satisfies Eq. 4.25 for $t \leq t_c$, then the choice of reserves satisfies $f(t) \geq f_{limit}$ for $t \leq t_c$. For an actual contingency, the frequency response might initially decrease faster than the linear approximation $f_{lin}(t) = f_{nom} - (f_{nom} - f_{limit})(t/t_c)$, this should cause a greater reserve response than the modeled response, $R_i(t) \geq R_{i,lin}(t)$ i.e. physical response greater than or equal to the modeled response with a linear frequency transient, as providers will see a greater urgency to respond. Since it is known that $R_{T,lin}(t)$ satisfies $f(t) \geq f_{limit}$, even more so should $R_T(t)$ satisfy it. Therefore the actual frequency will eventual rise above $f_{lin}(t)$ so that $f(t_c) \geq f_{limit}$.

A second condition is required to ensure that the $f(t) \geq f_{limit}$ is satisfied for all time, not just for $t \leq t_c$. This constraint is $R_T(t_c) \geq P_l$. This ensures the minimum frequency has occurred before t_c, when $\dot{f} = 0$ and $R_T = P_l$, and the frequency will no longer decline. A more rigorous proof is provided by [Sokoler et al. 2016].

At this stage Eq. 4.25 cannot be implemented into MILP as it is a function over continuous time domain. It is necessary to discretise the time at which the constraint is evaluated, for example Eq. 4.25 is evaluated at $t_k = 1, 2, 3, \cdots$ seconds. Further details can be found in the paper. At this point it is noticed that this approach is similar to that proposed by Miller [2014] at the Electricty Authority (EA) and his Area Under the Curve approach. Instead of using the linear frequency decline, Miller [2014] uses the IR test signal [SO 2016], which more closely approximates how the frequency will look in a transient. Therefore it is more accurate in approximating $f(t) \geq f_{limit}$. Wang et al. [2018] also discretises reserve responses and the frequency transient, but limits it to step changes in power output.

6. *Teng, 2016*

Teng et al. [2016] formulated a minimum frequency constraint that did not require linearisation, but the constraint is insufficient to ensure that the minimum frequency does not drop below the frequency limit. The formulation of the minimum frequency constraint starts by defining how reserve responds, as shown in Figure 4.4.

The standard swing equation is used to determine the evolution of frequency from an instantaneous loss of generation, ΔP_L in MW:

$$2H\frac{d\Delta f(t)}{dt} + D\Delta f(t) = \Delta P_g(t) - \Delta P_L \tag{4.30}$$

where $\Delta P_g(t)$ is the reserve response as shown in Figure 4.4, and is explicitly

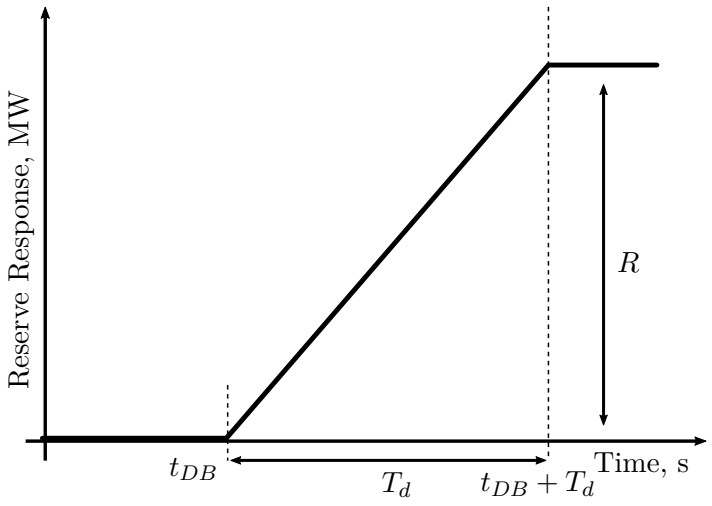


Figure 4.4 The model of reserve response used by [Teng et al. 2016] to formulate the minimum frequency constraint. The response starts after t_{DB}, once the frequency has dropped below the deadband frequency, Δf_{DB} which is a negative value, reserve starts to respond. The reserve response linearly increases over time T_d, to a reserve output of R MW, after this the reserve response stops.

defined in Eq. 4.31.

$$\Delta P_g(t) = \begin{cases} 0 & t < t_{DB} \\ \frac{R}{T_d}(t - t_{DB}) & t_{DB} \leq t < t_{DB} + T_d \\ R & t_{DB} + T_d \leq t \end{cases} \tag{4.31}$$

The units of the Inertia, H, are in MW s Hz^{-1}. Load damping, D, is in MW Hz^{-1} therefore the frequency deviation from the nominal frequency, Δf, is in Hz. Solving the differential, Eq. 4.30, for the first two time periods of Eq. 4.31, the frequency for the second time period is:

$$\Delta f(t) = \Delta f_{DB} - \left(\frac{\Delta P_L}{D} + \Delta f_{DB} + \frac{2HR}{T_d D^2}\right)\left(1 - e^{-\frac{D}{2H}t'}\right) + \frac{Rt'}{T_d D} \tag{4.32}$$

where $t' = t - t_{DB}$. A simplification is made where $\Delta P'_L = \Delta P_L + D\Delta f_{DB}$, and Δf_{DB} is here defined differently to [Teng et al. 2016]. The time of the minimum frequency is found by differentiating Eq. 4.32 with respect to time and solving for when the derivative is zero, the result is shown in Eq. 4.33. Then substituting the result back into Eq. 4.32, the minimum frequency is derived in Eq. 4.34.

$$t_{min} = t_{DB} - \frac{2H}{D} \ln\left(\frac{2HR}{T_d \Delta P'_L D + 2HR}\right) \tag{4.33}$$

$$\Delta f_{min} = \Delta f_{DB} - \frac{\Delta P'_L}{D} - \frac{2HR}{T_d D^2} \ln\left(\frac{2HR}{T_d \Delta P'_L D + 2HR}\right) \tag{4.34}$$

The minimum frequency constraint $\Delta f_{limit} \leq \Delta f_{min}$ is rearranged to obtain:

$$D^2(\Delta f_{limit} - \Delta f_{DB}) + D\Delta P'_L \leq \frac{2HR}{T_d D^2} \ln\left(1 + \frac{T_d \Delta P'_L D}{2HR}\right) \tag{4.35}$$

Teng et al. [2016] noticed that since the right-hand side of Eq. 4.35 is a monotonically increasing function of HR, the minimum frequency constraint can be reformulated as $HR \geq k$ where k is the solution to:

$$D^2(\Delta f_{limit} - \Delta f_{DB}) + D\Delta P'_L = \frac{2k}{T_d D^2} \ln\left(1 + \frac{T_d \Delta P'_L D}{2k}\right) \tag{4.36}$$

$HR \geq k$ is not yet a linear constraint as it is dependent on the combined total inertia and reserve. However, Teng et al. [2016], by using integer constraints, was able to formulate it into MILP, which can be found in his paper. $HR \geq k$ ensures that the minimum frequency constraint is satisfied if $t_{min} \leq t_{DB} + T_d$. To ensure $HR \geq k$ is the only minimum frequency constraint required, $t_{min} \leq t_{DB} + T_d$ needs to become a constraint in the formulation. It appears this constraint cannot be formulated into MILP, this can be seen by substituting Eq. 4.33 into $t_{min} \leq t_{DB} + T_d$ and rearranging to get:

$$\frac{2HR}{D} \ln\left(1 + \frac{T_d \Delta P'_L D}{2HR}\right) \leq RT_d \tag{4.37}$$

which describes the required relationship between HR and R. The trick that was used in Eq. 4.35 cannot be applied here precisely because R is separated from H on the right-hand side of the equation. Further research is needed to prove if Eq. 4.37 can be reformulated into a MILP format.

7. *Lee, 2013*

This paper includes the minimum frequency constraint in finding the optimal procurement of reserves and energy by multiple iterations of a master MILP problem. Compared to the previous papers where the optimal solution is found with one main MILP formulation, the iterative approach can reduce the computational burden that the minimum frequency constraint has when determining the optimal amount of reserve. Lee and Baldick [2013] provides a two stage process for finding the optimal solution between energy, reserve, and loss of load costs. The approach is quite technical, and to repeat the main ideas here would fully require repeating the paper. One comment is made, including the cost of lost load into the optimisation has significantly increased complexity, but to not consider its cost in reserve shortfall situations may under-value reserves and incorrectly incentivise capacity planning.

8. *Li, 2018*

   Li et al. [2018] developed a two product optimisation for contingency reserve. The approach closely follows that of Doherty et al. [2005], where linear constraints are derived by curve fitting results from transient simulations. The two products are Primary Frequency Response (PFR) from synchronous generation, and Fast Frequency Response (FFR) from interruptible load or energy storage. FFR is triggered 0.5 seconds after the frequency drops below 59.7 Hz. Transient simulations of ERCOT power system are conducted to determine the relative worth of PFR and FFR during different grid conditions, the result is a reserve requirement constraint:

$$\alpha P_{FFR} + P_{PFR} \geq FRR_{min} \tag{4.38}$$

   where P_{FFR} and P_{PFR} are reserve quantities, α is the relative worth, a dimensionless constant, and FFR_{min} is the minimum reserve requirement to cover the loss of two Nuclear units. The parameters α and FRR_{min} are adjusted before the optimisation depending on the amount of inertia available, to keep the frequency above 59.4 Hz and avoid load shedding at 59.3 Hz. Table 2 of [Li et al. 2018] shows how these parameters vary: α ranges from 2.2 to 1.0 when inertia is low at 239 GW s to when it is high at 593 GW s. A co-optimised energy and reserve simulation is performed. Liu and Du [2018] uses this method to perform an optimisation for a Day Ahead Market. Greve et al. [2018] have also found a method of comparing two different reserve products in Great Britian.

4.1.3 Steady State Frequency Constraint

1. *Restrepo, 2005*

   A steady state frequency constraint is concerned with the final frequency reached after a contingency. It is assumed that the final deviation is the result of free governor action (droop) to remove the imbalance between generation and supply. For this imbalance to be removed, steady state frequency has to be different than nominal, indicating to governors how much they should respond. However this frequency deviation may be too large. Therefore there are two options to correct this: the sensitivity of generation governors could be increased by decreasing the droop value, or more generation units with headroom could be dispatched. Changing the governor settings is an offline process; it is only the second option of increasing the number of online units which is considered in the dispatch process. Restrepo and Galiana [2005] provides the first significant work creating steady state frequency constraints, accounting for saturation in available reserve capacity, when a generator has no ability to increase power output.

To appreciate the contribution of [Restrepo and Galiana 2005], consider a simpler formulation of steady state frequency constraints. The droop characteristic of generators is in a percentage, 3 to 6%, which is translated to a gain in $\mathrm{MW\,Hz^{-1}}$, R_i, for each generator indexed by i. The relationship between frequency deviation from nominal and the change in governor output from the dispatch is:

$$\Delta P_i^R = -R_i \Delta f u_i \tag{4.39}$$

where u_i is a binary variable stating if the unit is in operation. The steady state frequency constraint limits the frequency deviation, $\Delta f \geq \Delta f_{min}$, while the total response has to equal the loss of generation, P^L:

$$P^L = \sum_i \Delta P_i^R \tag{4.40}$$

This simplified formulation has limited practicality as it does not consider the available capacity of generation. However, introducing a capacity constraint removes the generality of Eq. 4.39 and its linearity, as there becomes a frequency where the governor cannot increase output, as shown in Figure 4.5.

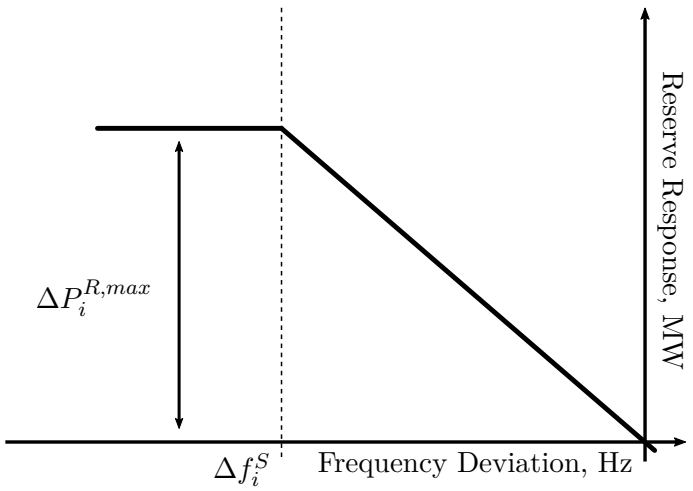


Figure 4.5 The steady state response of a generator governor to a frequency deviation. Once the frequency deviation drops below the saturation point, Δf_i^S, the generator has reached its capacity limit, $\Delta P_i^{R,max}$.

The key contribution of [Restrepo and Galiana 2005] is incorporating this relationship into a MILP form. A comment is made on two typographical mistakes in the paper: the second part of Eq. 16 should have a positive sign instead of a negative just after u_{it}. The first use of g_i^{max} in Eq. 17 and 18 should be r_i^{pr-max}.

2. *Ela, 2014*

Ela et al. [2014] ensured the steady state frequency does not exceed the limit by finding how much reserve through governor action would occur at the steady state frequency limit for each generator. Then the amount of reserve procured from each generator is limited below this value. This ensures that the amount of reserve needed, which is determine by the size of the contingency, is spread across enough units, creating a higher sensitivity to the load imbalance and keeping the steady state frequency within limit.

Ela et al. [2014] also includes consideration of a minimum frequency constraint and RoCoF constraint. To satisfy the minimum frequency constraint, dynamic simulations are run in sequences with the optimisations to update the parameters of the minimum frequency constraint, thereby creating an iterative approach. This approach is not too dissimilar to the interaction between SPD and RMT in New Zealand. Therefore, this work has not been reviewed in the previous section.

4.1.4 Further Research

There is further research looking at the dynamic effects on the power system and incorporating the minimum frequency constraint into the dispatch process, but have not explicitly formulated the frequency constraint or developed it further into an optimisation problem. This section explains these approaches, and the limitations recognised by the authors in incorporating these ideas into an electricity market formulation.

1. *Miller, 2016*

   Miller and Pajic [2016] is interested in valuing the response of inertia emulation from wind turbines. They developed a method to calculate the impact of inertia emulation on the minimum frequency during a transient. The method started by considering the impact of small quantity of finite duration reserve, like an impulse response, on the minimum frequency for a large transient on the Western Interconnection, using a detailed model. The quantity of reserve was 1000 MW for 0.5 seconds. The size of the contingency was the simultaneous trip of two units at the Palo Verde nuclear power station, 2750 MW. The benefit of the extra reserve was an increase in the minimum frequency, but the impact it had on increasing the minimum frequency was dependent on when the extra reserve was initiated. Miller and Pajic [2016] provides a plot of the relationship between the time at which the extra reserve is initiated and the impact it has on changing the minimum frequency in Figure 3 of his paper. The general relationship is also presented in Figure 4.6 here.

   From Figure 4.6, it can seen that the initiation time of the reserve has a large effect on the minimum frequency, and counter intuitive to what is expected, as it not the quickest response of reserve that gives the greatest impact. Rather it

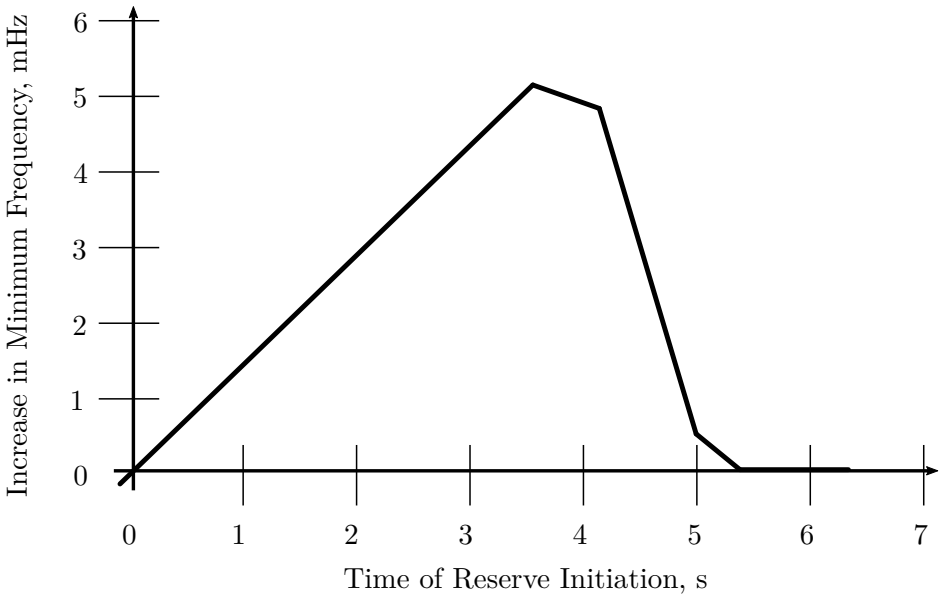


Figure 4.6 The impact of adding a short duration reserve to the minimum frequency for differences in initiation time. This Figure is reproduced from [Miller and Pajic 2016].

is when the end of the inertia emulation reserve is aligned with the minimum frequency, which is seen at about 3.5 seconds. The minimum frequency occurs just after 5 seconds for the unaided system. The reason that delaying the reserve provision has the better response is because it allows other reserve types to see the quickly dropping frequency and respond with greater impact. So having the inertia emulation response initiated earlier signals to other reserve types, via the frequency, that the response does not need to be as great, and the frequency drops lower.

The actual inertia emulation from wind turbines is more complicated than a short duration reserve output over 0.5 seconds, but as seen in Figure 5 of the paper it varies over time from 0 to 10 seconds, but to simulate a full power system during the dispatch process would not be practical. To avoid this simulation, the short duration reserve tests above could be used. The short duration tests are performed offline, and comprise impulse responses of the power system. If the total inertia emulation response ranging from 0 to 10 seconds is divided into 0.5s blocks, then impact of each 0.5 second block can be determined, and the summation of all blocks determines the total impact on the minimum frequency. It has been shown that this approach gives a good approximation.

2. *Transient Security Constrained Optimal Power Flow (TSCOPF)*

   There has been a lot of work in the literature developing techniques that incorporate transient stability constraints into the optimal power flow problem, hence the acronym TSCOPF. This is a summary of this work from the review papers of Abhyankar and Geng, [Abhyankar et al. 2017] and [Geng et al. 2017]. The problem is to minimise the cost of generation, or the real power losses on the network, etc. depending on the choice of generation. The power system is defined by a set of network equations, and models of synchronous machines. To ensure transient stability, constraints are placed on the allowable deviation of rotor angles for each generator. The problem is non-linear, and requires appropriate solvers. Abhyankar and Geng review the different means of solving these problems. One approach is to discretise the differential equations and avoid continuous optimal control techniques, but is limited by a how quickly the problem size escalates. Therefore other approaches have been used to simplifying the equations.

   The goal of this research is to develop problem formulations and optimisations to a level that can be used in the scheduling and dispatch process. However it is understood that the complexity of the problem is a major difficultly in arriving at this outcome, and the use of non-linear solvers brings a question of achieving global optimality and uniqueness of solution that is not helpful for a market implementation. However this work provides potential avenues for developing frequency constraints and solving potentially non-linear equations.

4.1.5 Suitability for New Zealand

This section evaluates the suitability of each method for improving the New Zealand IR market, through which the different methods are contrasted and organised. The suitability is evaluated on the following objective criteria:

1. Does the method optimise different reserve products, each defined by different response characteristics, e.g. time of delay after a contingency, instantaneous or gradual increase in power output? This criterion is important because it leverages the benefits of fast acting reserve in low inertia conditions, not having to simplify their performance to a common type in order to solve the optimisation problem, so that reserve is not over-procured in capacity constrained situations. Secondly it allows for potential pricing methodologies that can incentivise the right mix of resources.

2. Does the method fulfil the purposes of a co-optimised energy and reserves market from the New Zealand perspective? Is the contingency size optimised against the cost of reserve, and is generation capacity optimally divided between energy and reserve? It is not necessary to co-optimise in order to dispatch reserve, as they

can be optimised separately, but co-optimisation better incentivises generation capacity investment.

3. Does the method accurately predict the frequency transient for a dispatch of reserves, and is this done in a computationally appropriate time? That is can the optimisation be solved in the one to two minute time range, allowing for dispatches every five minutes. Two issues are of concern here: accuracy means that not too little or too much reserve is procured, avoiding an insecure power system or an expensive one, but accuracy comes at cost of computation time. This criterion is quite subjective.

4. Does the method procure enough reserve for the frequency to return to nominal? If only enough reserve was procured to cover the largest credible contingency, then frequency will stop declining but it will not return to nominal either, therefore more reserve is needed. New Zealand currently does this through the RMT, but few methods explicitly consider this even though it may be included implicitly.

5. Does the method assume the response characteristics scale in proportion to the amount of capacity provided? For example, is the rate at which reserve increases in proportion to how much is dispatched? If 60 MW is procured and reserve ramps at a rate of 10 $\mathrm{MW\,s^{-1}}$, what happens when 30 MW is procured? Does the ramp rate remain at 10 $\mathrm{MW\,s^{-1}}$ or does it drop to 5 $\mathrm{MW\,s^{-1}}$. This may seem a strange question, as the response always follows the controller set by the power station, but in the process of linearisation this may be lost. The concern is dynamic capabilities should not be lost on account of the amount dispatched, unless there is a physical reason.

Not every method is going to fulfil all of these criteria, although some are close. The point is that there is not a perfect option in the literature, and so there is potential for improvement. To avoid comparing every option in equal detail, criterion 1 is used to divide them into three categories: a single reserve product, a finite number of products, or an indefinite number. Methods with greatest generality in the third category are analysed in more detail. The methods are organised as follows:

- Single Reserve Product
  - [Doherty et al. 2005]
  - [Ahmadi and Ghasemi 2014]
  - [Teng et al. 2016]
- Finite Number of Reserve Products
  - [Li et al. 2018] and [Liu and Du 2018]

Table 4.1 Assessment of the four main methods against the criteria for suitability in New Zealand.

	Method			
Criteria	Lee, 2013	Wen, 2016	Sokoler, 2016	Wang, 2018
1	Constant ramping reserve products, ramp rate defined, initiated at time of contingency.	Constant ramping reserve products, ramp rate defined, initiated at time of frequency leaving deadband, and instantaneous battery response.	Any possible response, tested with linear frequency signal.	Instantaneous reserve output, defined by delay after contingency.
2	Yes	Yes	Yes	Yes
3	Two stage iterative optimisation, for the purpose of solve speed. Linearisation creates uncertianity in solution optimality.	Fast solving. Convexity of non-linear constraint unknown, linearisation may not direct towards finding the global minimum.	Solve speed and accuracy traded between by discretisation sampling time.	Solve speed and accuracy traded between by discretisation sampling time.
4	No	No	No	No
5	No, ramp rate invariant to quantity dispatched.	No, ramp rate invariant to quantity dispatched.	Yes, response scaled in proportion to quantity dispatched.	No, instantaneous responses are invariant to quantity dispatched.

  - [Greve et al. 2018]

- Indefinite Number of Reserve Products
  - [Lee and Baldick 2013]
  - [Chávez et al. 2014]
  - [Wen et al. 2016]
  - [Sokoler et al. 2016]
  - [Wang et al. 2018]

Of the five methods that remain, the second criterion removes [Chávez et al. 2014] from consideration, because including contingency size into the formulation makes the equations non-linear. The effort of Chávez et al. [2014] continues with Wen et al. [2016]. Secondly, [Wang et al. 2018] remains in consideration, as the formulation can be easily applied in a co-optimised problem, but is not made explicit. Hence four methods are left, and are briefly compared against each criterion in Table 4.1.

Not counting criterion four, there is a clear difficulty in satisfying all the other criteria simultaneously. The first and fifth criteria appear antithetical: [Lee and Baldick

2013], [Wen et al. 2016], and [Wang et al. 2018], which do not have the same generality in reserve products as [Sokoler et al. 2016], all satisfy criterion five which [Sokoler et al. 2016] does not. There is a clear intractability between having multiple different reserve products and defining a solvable minimum frequency constraint for it. Even for simpler portfolios of reserve products a linearisation has to be introduced. When generality is proposed it comes at the cost of having to distort the reserve response away from its maximum capability, and degrading the accuracy of the minimum frequency constraint.

This intractability of formulating a minimum frequency constraint may seem impossible to avoid, but it provides the motivation for this thesis. In the next chapter a new methodology is developed that can satisfy all five criteria, removing the intractability and creating a practical formulation that could be used to improve New Zealand's reserve market. The new methodology further generalises the instantaneous and constant ramping products, similar to those of [Wen et al. 2016], and avoids any linearisation process or discretisation. However, there is a cost, complexity, which may limit its adoption.

4.2 OPTIMAL CONTROL

From describing the problem so far, i.e. optimising control responses to satisfy the constraints of a dynamic system, the problem sounds very similar to the area of research called Optimal Control. The size of this research field is large, e.g. the *SIAM Journal on Control and Optimization* is dedicated to it, and in several other journals it features highly. Secondly there are a large number of textbooks with application to Economics and Engineering. However, the formulation presented in the next chapter avoids Optimal Control for two reasons: the objective functions are different, and translating the problem from Optimal Control to economic dispatch formulations is too difficult. The rest of this section briefly describes Optimal Control to elaborate on these reasons, and an important insight is highlighted.

The standard continuous control problem is to minimise:

$$\phi(\mathbf{x}(T), T) + \int_{t_0}^{T} L(\mathbf{x}(t), \mathbf{u}(t), t)dt \tag{4.41}$$

subject to the dynamic equations of the system:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{4.42}$$

where $\mathbf{x}(t)$ are the state variables, and $\mathbf{u}(t)$ are the control variables, which are both functions of time, t. For example, the grid frequency could be a state variable and the reserve responses are the control variables. The objective is defined by two cost functions, ϕ and L, the former dependent on the final condition, the later integrated over the time of interest. Therefore what is being optimised is not individual variables, but rather functions, and hence KKT conditions cannot be used to solve this problem, but rather results from the Calculus of Variations (Elsgolc [2012] provides a good introduction to this subject). The full optimal conditions can be found in any textbook on the subject, e.g. Lewis et al. [1995], but to simplify the conditions it is helpful to define the Hamiltonian:

$$H(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t), t) = L(\mathbf{x}(t), \mathbf{u}(t), t) + \boldsymbol{\lambda}(t)^T \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{4.43}$$

The optimal control sequence, $\mathbf{u}(t)$, is found when:

$$\frac{dH}{dt} = 0 \tag{4.44}$$

provided that the Hamiltonian is independent of time, $H = H(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t))$. If the Hamiltonian is dependent on time more terms are required.

The difference between optimising variables and functions is the main justification for not considering Optimal Control further. The primary goal for reserve optimisation

is to determine how much capacity should be made available for reserve to balance a contingency, therefore $\max\{\mathbf{u}(t)\}$ is of interest, not the integral of $\mathbf{u}(t)$. I am unaware of methods directing interest to the maximum, other than that of Discrete Optimal Control, which [Sokoler et al. 2016], [Miller 2014], and [Wang et al. 2018] are unconventional examples. To translate the problem into a conventional Optimal Control form is too difficult and unnecessary given the result of the next chapter.

Inequality constraints can be placed on the control variables, $\mathbf{g}(\mathbf{u}(t), t) \leq 0$. The process of finding the optimal solution is derived from Pontryagin's Minimum Principle on the Hamiltonian:

$$H(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*, t) \leq H(\mathbf{x}^*, \mathbf{u}, \boldsymbol{\lambda}^*, t) \quad \text{for all feasible } \mathbf{u} \tag{4.45}$$

The consequence is $\mathbf{u}(t)$ satisfies $g_i(\mathbf{u}(t), t) = 0$ for some constraints, i, when the Hamiltonian derivative in Eq. 4.44 cannot be zero, and so the constraints g_i define the optimal control response. This effect results in what is called Bang-Bang control, as $\mathbf{u}(t)$ swings between the limits defined by the inequality constraints. Relating this to the problem of grid frequency and reserve response, only the maximum capability of reserve providers is of primary importance. For example, in a contingency the wicket gates of a hydro power station should be opened as quickly as possible, to provide the most power, but limited to not damaging equipment or extending past the maximum generation capabilities. Therefore transfer functions of the control and governor systems can be omitted from the problem, as long as reserve providers can suspend normal controls in contingencies. They then should be re-engaged when the frequency has returned to a safe range, thereby ensuring long term stability. The formulation of the next chapter assumes this is what happens for SR, as suspension of normal control in these situations is not an uncommon practice in New Zealand.

For a practical problem, limits are required on grid frequency, a state variable. Pontryagin's Minimum Principle can not be applied directly in these situations, and the assumption of Bang-Bang control loses validity. A significant amount of research is available on these constraints, Hartl et al. [1995] provides a review of this analysis, but may be out of date. Nonetheless it is assumed that Bang-Bang control is an appropriate control strategy, because when grid frequency falls quickly, all should be done to arrest it.

Chapter 5

OPTIMISATION FORMULATION

In the previous chapter means of incorporating frequency dynamics into the reserve optimisation is reviewed. In this chapter a new approach of optimising reserve is presented. This method which is unconstrained to linearisation techniques, and uses a sequential algorithm to move from one feasible solution region to the next in order to find the optimal solution. Each region is defined by linear and quadratic constraints that can be solved by a Quadratically Constrained Programming (QCP) solver. The entire feasible solution space still retains convexity like Linear Programming (LP) and shares similar solution uniqueness properties. Therefore a significant effort is put into proving convexity.

5.1 PROBLEM DESCRIPTION

The main motivation for recognising the finite response time of reserve providers is that if such a consideration was not made, then the reserve dispatch may not achieve its objective of keeping the frequency above its minimum limit, i.e. the choice of reserve acts too slowly and by the time reserve has covered the risk the frequency has dropped too far. However, the New Zealand system is aware of this concern and dispatches Fast Instantaneous Reserve (FIR) above what the risk requires, to increase the response rate. The extra FIR capacity is not utilised, presenting an inefficiency in the market system. This is most felt when capacity is at a premium, as highlighted by the potential future of the New Zealand power system in Chapter 3. For other systems, which do not have the ability to dynamically change reserve requirements across schedules, the consequences are worse as the power system collapses and blackouts can ensue.

The second motivation is to increase competition in the reserve market and thereby increase efficiency. For example, Instantaneous Reserve (IR) products in New Zealand are made of three different types: Interruptible Load (IL), Partially Loaded Spinning Reserve (PLSR), and Tail Water Depressed Reserve (TWDR). The last two are referred to as Spinning Reserve (SR). IL capability is defined by how much load can be disconnected one second after the frequency drops below 49.2 Hz. Some IL providers cannot shed

load in one second but are close and yet cannot offer into the FIR market. Moving to a market where response time is considered increases the participation of these providers, removing a barrier to entry.

In Section 2.6 new providers of reserve are presented, these do not necessarily follow current definitions, neither would fitting them to current definitions fully value the service they provide. Letting each reserve provider be defined by its own characteristics would be ideal.

The requirements for a new reserve market are:

1. A general reserves option not just defined by quantity and price, but response time as well.
2. A means of relating these reserve options to the frequency transient, and limiting the frequency to be within constraints.
3. A set of constraints that can be co-optimised with an energy market.

A pricing methodology is required for a complete reserves market implementation, but it is not developed in this work. The standard method of having the reserve price determined by the Lagrangian of the reserve requirement constraint is no longer appropriate for this optimisation, because the Lagrangian may be smaller than some of the reserve prices that have been dispatched. Also it is not desired that each reserve provider have the same price as each is a slightly different service. Therefore due to time constraints a pricing methodology is not developed.

Motivation for this research can also be found in international power systems. The benefit of this new reserve formulation is that it can simplify reserve markets for contingencies. For example, in the power system of Great Britain, there is a reserve category for very fast responses within one second called Enhanced Frequency Response (EFR), and a reserve category for providers that respond around ten seconds called Primary Response. The new formulation can combine the optimisation of these two classes of reserves by the formulation of a generalised reserve option.

The goal of this chapter is to formulate an optimisation problem that has the potential to be co-optimised with any energy or capacity market around the world. More specifically any market that already utilises a MILP solver or Mixed Integer Quadratically Constrained Programming (MIQCP) solver in the procurement of reserves. This chapter does not show a co-optimised problem, as further development is required for the inclusion of mixed integer variables into the optimization. However 'potential' is defined by the formulation being convex, and the existence of a practical solver. Therefore this chapter is broken into three main sections after the problem description: demonstrating the convexity of the problem, the development of a solver, and evaluating the performance of the solver. The last major section also highlights specific consequences of applying this formulation.

In this chapter, the main form of the optimisation considered is to:

minimise *reserve availability cost*
subject to *frequency constraints*
and reserve availability limits

where the objective is to minimise a linear function of continuous variables that are the amount of reserve capacity made available. In a real-time market, this will be amount of reserve dispatched in MW. The cost is optimised within the constraints of keeping the power system frequency above the limit, and bounds on the variables. For a full implementation of a reserves optimisation, more constraints will be required, but to keep the generality of the approach to international markets, only the core constraints to this new formulation are given.

It will be shown that the form of the optimisation problem cannot be expressed by a single set of equality and inequality constraints, as the form of the constraints varies depending on location in the feasible solution space. Therefore the problem cannot be classified as QCP. Instead the feasible solution is divided into regions which are generalised and are classified as QCP. This class of problem is given the name Piecewise Quadratically Constrained Programming (PQCP) after the piecewise nature of the frequency constraints.

5.1.1 The Reserve Option

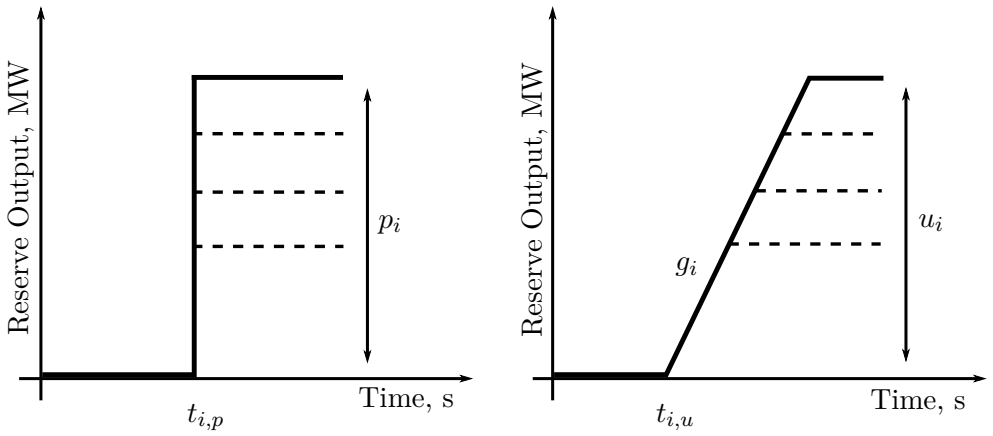
The Reserve Option comes in two different types: an instantaneous increase in reserve output, called the IL type; and the finite response time, called the SR type. Each option is a function of time, and is adjusted by their reserve dispatched parameter, p_i and u_i respectively. The function for the IL type is:

$$P_i(t; p_i) = \begin{cases} 0 & t < t_{i,p} \\ p_i & t \geq t_{i,p} \end{cases} \tag{5.1}$$

Each option is indexed by i. The time at which reserve response instantaneously increases to p_i is the IL trigger time $t_{i,p}$, as seen in Figure 5.1a. The function for the SR type is shown in Eq. 5.2 and in Figure 5.1b:

$$U_i(t; u_i) = \begin{cases} 0 & t < t_{i,u} \\ g_i(t - t_{i,u}) & t_{i,u} \leq t < t_{i,u} + u_i/g_i \\ u_i & t_{i,u} + u_i/g_i \leq t \end{cases} \tag{5.2}$$

Once SR is initiated at $t_{i,u}$, the reserve output ramps up at a rate of g_i MW s^{-1}, eventually reaching the reserve dispatch, u_i. The units of reserve quantities is in MW and for time seconds. The IL option is seen as a special case of the SR type with



(a) $P_i(t; p_i)$ (b) $U_i(t; u_i)$

Figure 5.1 The profile of IL and SR offers. The dotted lines present possible changes in the dispatch with reducing output.

an infinite ramp rate. Therefore there is only one general reserve option, but for the purposes of mathematical rigour a distinction has to be made.

5.1.2 The Frequency Transient

The change in Power System frequency is dependent on the balance of mechanical power provided and electrical load, the relationship is coupled by the inertial constant, H in MW s:

$$2H\frac{df(t)}{dt} = P_M(t) - P_E(t) \tag{5.3}$$

The frequency, $f(t)$, is in per unit with a base of 50 Hz and is a perturbation from 1 pu. For the largest credible contingency there is a potential instantaneous loss of generation, R, and is referred to as risk. The reserve responds and the mechanical and electrical balance becomes:

$$2H\frac{df(t)}{dt} = -R + \sum_i P_i(t; p_i) + \sum_i U_i(t; u_i) \tag{5.4}$$

To avoid the tripping of more generation the frequency has to be kept above the

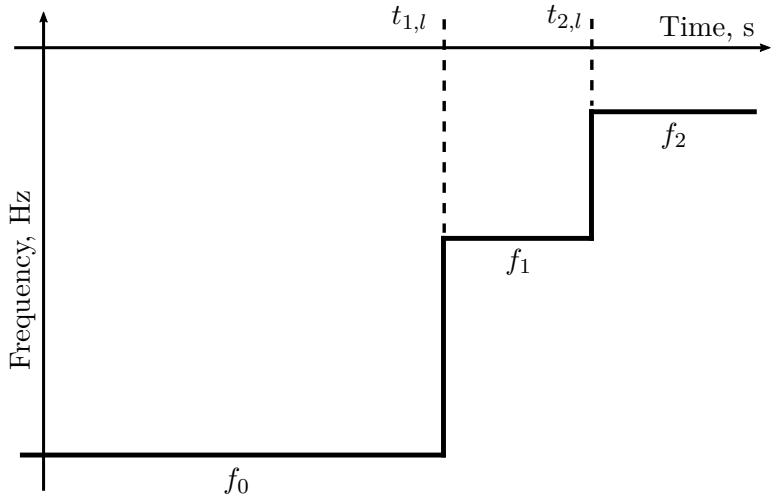


Figure 5.2 The frequency limit function, the power system frequency has to be above this limit for a contingency.

limit, $f(t) \geq f_{lim}(t)$, as presented in Figure 5.2, and defined by:

$$f_{lim}(t) = \begin{cases} f_0 & 0 \leq t < t_{1,l} \\ f_1 & t_{1,l} \leq t < t_{2,l} \\ \vdots & \\ f_j & t_{j,l} \leq t < t_{j+1,l} \\ f_{N_c} & t_{N_c,l} \leq t \end{cases} \tag{5.5}$$

where $t_{j,l}$ determines the time of each step change in frequency limit from f_{j-1} to f_j. A given frequency limit function $f_{lim}(t)$ has Nc step changes. The frequency limit function is strictly increasing at each step change, $f_j > f_{j-1}$.

The frequency transient, $f(t)$, is found by integrating Eq. 5.4, with the initial frequency being zero, $f(0) = 0$:

$$\begin{aligned} 2Hf(t) &= \int_0^t -R + \sum_i P_i(\tau; p_i) + \sum_i U_i(\tau; u_i) d\tau + f(0) \\ &= -Rt + \sum_{t_{i,p} \leq t} p_i(t - t_{i,p}) + \sum_{t_{i,u} \leq t < t_{i,e}} (U_i(t; u_i)(t - t_{i,u}) - \frac{U_i(t; u_i)^2}{2g_i}) \\ &+ \sum_{t_{i,e} \leq t} (u_i(t - t_{i,u}) - \frac{u_i^2}{2g_i}) \end{aligned} \tag{5.6}$$

where $t_{i,e} = t_{i,u} + u_i/g_i$, the time at which an SR option stops ramping. The goal is to transform Eq. 5.6 into a series of finite conditions that satisfy $f(t) \geq f_{lim}(t)$.

Condition 1 - Reserve Requirement

The first condition requires that the frequency transient should stop falling, there has to be a time when the amount of reserve equals the risk:

$$-R + \sum_i p_i + \sum_i u_i \geq 0 \tag{5.7}$$

Condition 2 - Frequency Limits

At the time of each transition from one frequency level to another, $t_{j,l}$, the frequency has to be above f_j. Therefore defining $u_{i,j} = U_i(t_{j,l}; u_i)$, the Nc frequency limits are:

$$2Hf_j \leq -Rt_{j,l} + \sum_{t_{i,p} \leq t_{j,l}} p_i(t_{j,l} - t_{i,p}) + \sum_{t_{i,u} \leq t_{j,l}} (u_{i,j}(t_{j,l} - t_{i,u}) - \frac{u_{i,j}^2}{2g_i}) \tag{5.8}$$

Condition 3 - Minimum Frequency Constraint

Condition 2 satisfies the requirement that $f(t) \geq f_{lim}(t)$ for a finite number of points, $t_{j,l}$, but to satisfy for all points in time, one more constraint is required, the minimum frequency constraint. This constraint requires at the time of minimum frequency, i.e. when $df/dt = 0$, the frequency transient has to be above the frequency limit function. Before this constraint is expressed, the first time when minimum frequency is reached, t_{min}, is defined by the following conditions:

$$-R + \sum P_i(t_{min}; p_i) + \sum U_i(t_{min}; u_i) \geq 0 \tag{5.9}$$

$$\text{and} \quad \forall t < t_{min} \quad -R + \sum P_i(t; p_i) + \sum U_i(t; u_i) < 0 \tag{5.10}$$

The reason why t_{min} is defined this way is because there may be no time when $df/dt = 0$ as IL offers can instantly increase the reserve output, i.e. the right hand side of Eq. 5.4. Therefore Eq. 5.9 ensures the minimum frequency has occurred so that $df(t_{min})/dt \geq 0$. Then 5.10 ensures that t_{min} is earliest possible value for t_{min} by having all the values before the time t_{min} resulting in $df/dt < 0$. Eq. 5.10 also uniquely defines t_{min} in the situation when $df/dt = 0$ for a period of time, and determines t_{min} to be first time in this period. The minimum frequency constraint is found by substituting

$t = t_{min}$ into Eq. 5.6:

$$2Hf_{lim}(t_{min}) \leq -Rt_{min} + \sum_{t_{i,p} \leq t_{min}} p_i(t_{min} - t_{i,p}) + \sum_{t_{i,u} \leq t_{min}} (u_{i,m}(t_{min} - t_{i,u}) - \frac{u_{i,m}^2}{2g_i}) \tag{5.11}$$

where $u_{i,m} = U_i(t_{min}; u_i)$. For this to be the only other frequency constraint required, it has to be shown for all time that $f(t) \geq f_{lim}(t)$. For $t < t_{min}$ the frequency is always decreasing, and since f_j is strictly increasing then $f(t) \geq f_{lim}(t)$. For $t > t_{min}$ assuming Condition 2 is satisfied, since $df/dt \geq 0$ after t_{min} as the amount of reserve can only increase, it is not possible for $f(t) < f_{lim}(t)$ as that would require $df/dt < 0$ at some point in time after t_{min}.

Objective and Reserve Limits

To complete the optimisation problem it is necessary to have an objective function, 5.12, which is linear; the prices $c_{i,p}$ and $c_{i,u}$ are in units of \$/MW and can be negative, zero, or positive.

$$\sum c_{i,p} p_i + \sum c_{i,u} u_i \tag{5.12}$$

It is also necessary to set limits to how much reserve can be dispatched.

$$0 \leq p_i \leq p_i^{max} \tag{5.13}$$

$$0 \leq u_i \leq u_i^{max} \tag{5.14}$$

It is not possible for p_i and u_i to be below zero, as this would require more frequency constraints not expressed in Condition 1-3. Moreover it would negate the analysis used to prove convexity of the feasible solution space, and add complexity to the solving methodology that has not been included in this thesis. It is left to further research in Chapter 6 to discuss the potential of negative reserve responses in order to approximate synthetic inertia type responses from wind turbines.

5.1.3 Modelling Assumptions

The formation of reserve options and frequency transients have assumed some simplifications. This section explains some of these assumptions, and are listed as follows:

1. In the managing of grid frequency during contingencies, long term stability is ensured by the feedback that exists between power output of reserve providers

and the grid frequency. Generators respond to deviations in grid frequency from nominal, and then those real power output changes influence the rotational energy on the power system and the grid frequency. This feedback loop is broken in this formulation, where it is assumed reserve responses do not change for difference in the frequency transient.

The breaking of this feedback loop means that long term frequency stability is not guaranteed by this formulation. Another constraint could be added to ensure enough reserve providers with droop based controls are dispatched, but this has not been analysed. To determine the values of $t_{i,p}$, $t_{i,u}$, and g_i, the reserve response of a provider has to be tested against a standard frequency transient, like what is already done for testing the capabilities of current reserve providers [SO 2016].

2. It is assumed that all positive increases in reserve response can be modelled by a delay and either an instant or ramping increase in power output. The reserve response from load tripping once a certain frequency has been reached can be accurately modelled by an instantaneous response (IL offer). However, the reserve response from hydro and thermal generation may not be accurately modelled by a single ramped output, as these responses vary in slope. It is possible to model a single reserve provider with multiple SR offers, by linking the time one offer finishes ramping with the time next offer starts ramping, and requiring the price of each offer to increase in sequence. Also the prices can be separated from individual offers and associated with the combined output with extra equality constraints. With more SR offers, the better the response can be modelled.

3. The differential equation that determines the frequency transient, Eq. 5.4, may give the impression that natural responses of loads and the responses of generators that do not offer into a reserve market do not contribute to mitigating power imbalances. These responses can be approximated by IL and SR offers with zero price, to account for the impact these responses have on the frequency transient.

4. It is assumed that this formulation is only valid in modelling the frequency transient in the first frequency swing, from the initial decline to the rebound back to nominal frequency. An appropriate frequency limit function, $f_{lim}(t)$, is required to ensure this rebound, as analysed in Section 5.4.2. However, to model the frequency transient past the time the frequency returns to nominal, would require negative response offers, which have not been included.

5.2 CONVEXITY

The goal of this section is to show the space defined by the vector:

$$\mathbf{v} = [H, R, p_1, \cdots, p_{Np}, u_1, \cdots, u_{Nu}]^T \in X \subset \mathbb{R}^{2+Np+Nu}$$

bounded by the constraints Eq. 5.7, 5.8, 5.11, 5.13, and 5.14 is convex. Np and Nu are the total number of IL and SR offers respectively. The method of proving convexity for a space X is to show that X conforms with the definition of convexity, i.e. for every two elements $\mathbf{x}_1, \mathbf{x}_2 \in X$ then each point in between is also an element of X, or more precisely:

$$\forall \lambda \in [0, 1] \quad (1 - \lambda)\mathbf{x}_1 + \lambda \mathbf{x}_2 \in X \tag{5.15}$$

where λ is a parameter that transcribes a line between $\mathbf{x}_1$ and $\mathbf{x}_2$.

For example, consider the non-convex space of Figure 5.3. The straight line between points $\mathbf{x}_1$ and $\mathbf{x}_2$ is not entirely contained within the space, therefore the space is not convex. However for the convex space, it is seen that any line that starts and ends within the space is also contained in it entirely. The benefit of optimising over a convex space with a linear objective function is that there is unique solution on the boundary of the space. The only situation when there is not a unique solution is when the optimal boundary is flat, and the cost does not change along it. Therefore when it comes to solving the problem, only one optimal solution needs to be found, and is computational faster than having to find multiple minimums of a non-convex space.

The approach to proving the convexity of the problem space is progressive. It is obvious that the space defined by Eq. 5.13 and 5.14 is convex as it is just a higher dimensional cube. Then the first proof is to add Eq. 5.7 for the purposes of showing a simpler example before adding the frequency constraints, Eq. 5.8. Finally the more difficult proof is showing the space bound by Eq. 5.11 is convex.

5.2.1 Reserve Requirement Constraint

The region created by a plane such as Eq. 5.7 is convex as it is linear, but to show this through analysis, is to start with any two points in the space:

$$\mathbf{v}_1 = [H_1, R_1, p_{1,1}, p_{2,1} \cdots, p_{Np,1}, u_{1,1}, u_{2,1}, \cdots, u_{Nu,1}]^T$$

$$\mathbf{v}_2 = [H_2, R_2, p_{1,2}, p_{2,2} \cdots, p_{Np,2}, u_{1,2}, u_{2,2}, \cdots, u_{Nu,2}]^T$$

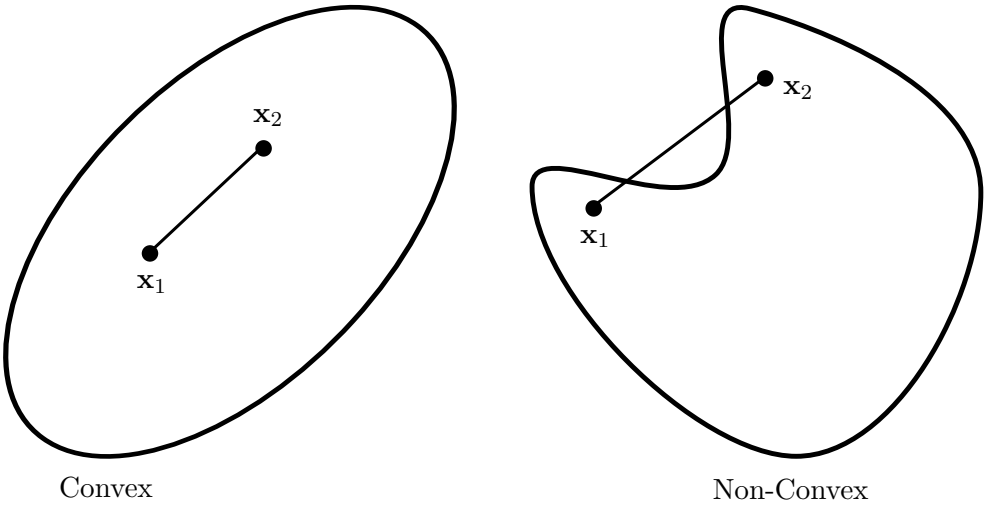


Figure 5.3 Examples of two dimensional space both convex and non-convex.

that satisfy the reserve requirement constraint of Eq. 5.7

$$-R_1 + \sum_i p_{i,1} + \sum_i u_{i,1} \geq 0 \tag{5.16}$$

$$-R_2 + \sum_i p_{i,2} + \sum_i u_{i,2} \geq 0 \tag{5.17}$$

To prove the convexity of the space is to show that $\mathbf{v} = (1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2$ also satisfies Eq. 5.7:

$$= -(1-\lambda)R_1 - \lambda R_2 + \sum_i((1-\lambda)p_{i,1} + \lambda p_{i,2}) + \sum_i((1-\lambda)u_{i,1} + \lambda u_{i,2}) \tag{5.18}$$

$$= (1-\lambda)(-R_1 + \sum_i p_{i,1} + \sum_i u_{i,1}) + \lambda(-R_2 + \sum_i p_{i,2} + \sum_i u_{i,2}) \geq 0 \tag{5.19}$$

which is greater than or equal to zero because $\lambda \in [0, 1]$.

5.2.2 Frequency Limits

The convexity of each frequency limit (Eq. 5.8) is shown individually at each time point, $t_{j,l}$, then the intersection of all frequency limit spaces is shown to be convex. It is helpful to visualise the feasible region of a single frequency limit. Consider a problem where the inertia and risk are constant and there are two SR offers with variables, u_1 and u_2. There is also only one step change in the frequency limit function. The constraint for that step change, Eq. 5.8, is simplified to:

$$2Hf_1 \leq -Rt_{1,l} + \left(u_{1,1}(t_{1,l} - t_{1,u}) - \frac{u_{1,1}^2}{2g_1}\right) + \left(u_{2,1}(t_{1,l} - t_{2,u}) - \frac{u_{2,1}^2}{2g_2}\right) \tag{5.20}$$

rearranging to this form:

$$-2Hf_1 - Rt_{1,l} + \frac{1}{2}g_1(t_{1,l} - t_{1,u})^2 + \frac{1}{2}g_2(t_{1,l} - t_{2,u})^2 \geq \frac{\left(u_{1,1} - g_1(t_{1,l} - t_{1,u})\right)^2}{2g_1} + \frac{\left(u_{2,1} - g_2(t_{1,l} - t_{2,u})\right)^2}{2g_2} \tag{5.21}$$

where $u_{1,1} = U_1(t_{1,l}; u_1)$ and $u_{2,1} = U_2(t_{1,l}; u_2)$, note for this example that this use of indices is different than the previous section. The feasible space is partly defined by an ellipse, centred at the coordinates $(g_1(t_{1,l} - t_{1,u}), g_2(t_{1,l} - t_{2,u}))$, as seen by an elliptical curve in its lower left quadrant in Figure 5.4. Outside of the lower left quadrant of the ellipse to the right, when $u_1 > g_1(t_{1,l} - t_{1,u})$, $u_{1,1}$ does not increase any more as any extra reserve is offered after $t_{1,l}$ does not count in satisfying the constraint, and the ellipse transitions to a straight line in the lower right quadrant of Figure 5.4. From Figure 5.4 it is clear that the region defined by the single frequency limit is convex. The next step is to generalise this result.

There are two options for $u_{i,j} = U_i(t_{j,l}; u_i)$: either $U_i(t; u_i)$ stops ramping before $t_{j,l}$ ($t_{i,u} + u_i/g_i < t_{j,l}$), or continues ramping after $t_{j,l}$ ($t_{i,u} + u_i/g_i \geq t_{j,l}$). Therefore $u_{i,j}$ is equal to

$$u_{i,j} = \begin{cases} u_i & t_{i,u} + u_i/g_i < t_{j,l} \\ g_i(t_{j,l} - t_{i,u}) & t_{i,u} + u_i/g_i \geq t_{j,l} \end{cases} \tag{5.22}$$

For ease of notation, define $w_{i,j} = g_i(t_{j,l} - t_{i,u})$. The consequence of have two forms for $u_{i,j}$ on Eq. 5.8 is that there are two different equations for Eq. 5.8. Since there are two SR offers for the current example, there are a total of four constraints. The forth is not visualised in Figure 5.4 as that constraint does not include u_1 or u_2 as variables. If there are Nu different spinning reserve offers that can dispatch past $t_{j,l}$ then there are 2^{Nu} possible constraints. When $u_i = w_{i,j}$ defines a boundary between when each equation should be applied, also the boundary between quadrants in the ellipse, and is

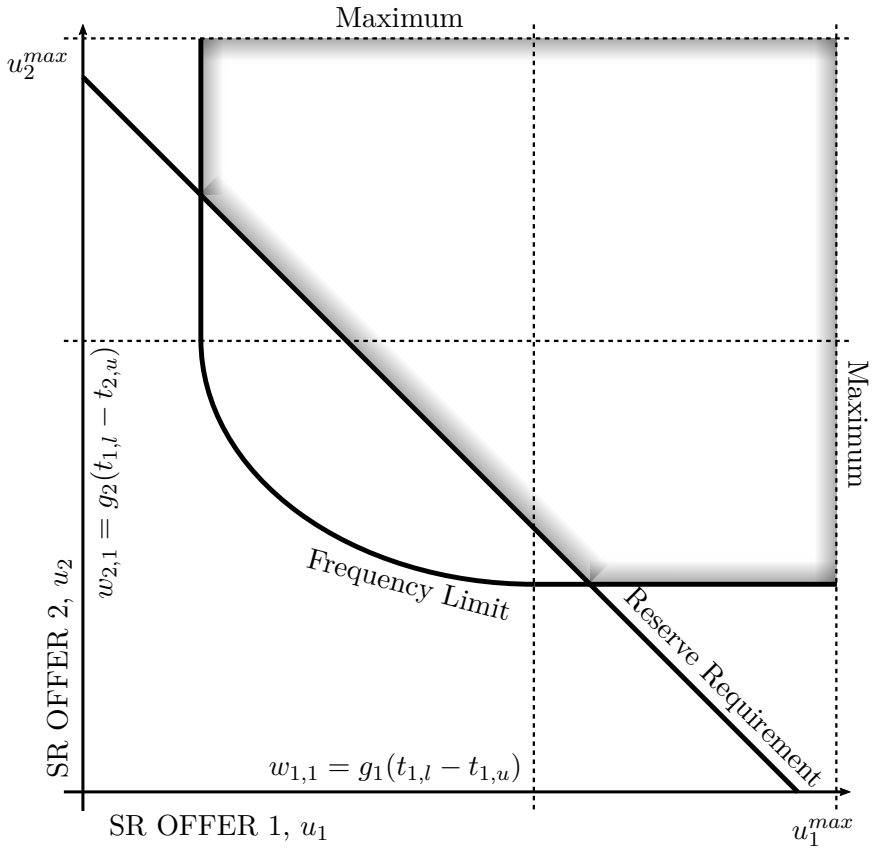


Figure 5.4 The feasible space, shaded, for a two SR offer problem with one step change in frequency limit at $t_{1,l}$. The feasible space is at the moment only defined by the reserve requirement constraint and the frequency limit.

seen by the internal horizontal and vertical dotted lines of Figure 5.4. There is one set of regions for $u_i \leq w_{i,j}$ and another for $u_i \geq w_{i,j}$.

To minimise the number of frequency limit constraint forms to one in proving convexity, a transformation is used to consider only what is dispatched before $t_{j,l}$, as any reserve that comes from ramping after $t_{j,l}$ has no influence on whether the frequency limit is satisfied or not. This transformation is symbolised as follows:

$$\mathbf{w} = T_j(\mathbf{v}) \tag{5.23}$$

where $\mathbf{w} = [H, R, q_1, q_2, \cdots, q_{Np}, w_1, w_2, \cdots, w_{Nu}]$. The transformation rules require that inertia, H, and risk, R, remain the same between $\mathbf{v}$ and $\mathbf{w}$. The rules for IL and SR offers are:

$$q_i = \begin{cases} p_i & t_{i,p} \leq t_{j,l} \\ 0 & t_{i,p} > t_{j,l} \end{cases} \tag{5.24}$$

$$w_i = \begin{cases} u_i & t_{i,u} + u_i/g_i < t_{j,l} \\ w_{i,j} & t_{i,u} + u_i/g_i \geq t_{j,l} \quad \text{and} \quad t_{i,u} \leq t_{j,l} \\ 0 & t_{i,u} > t_{j,l} \end{cases} \tag{5.25}$$

What this transformation looks like on the feasible space is shown in Figure 5.5. Notice all the points compress into the bottom left quadrant, that of the elliptic constraint. Therefore one form of the frequency limit constraint is required to assess if the dispatch satisfies the frequency constraint once the transformation has been applied.

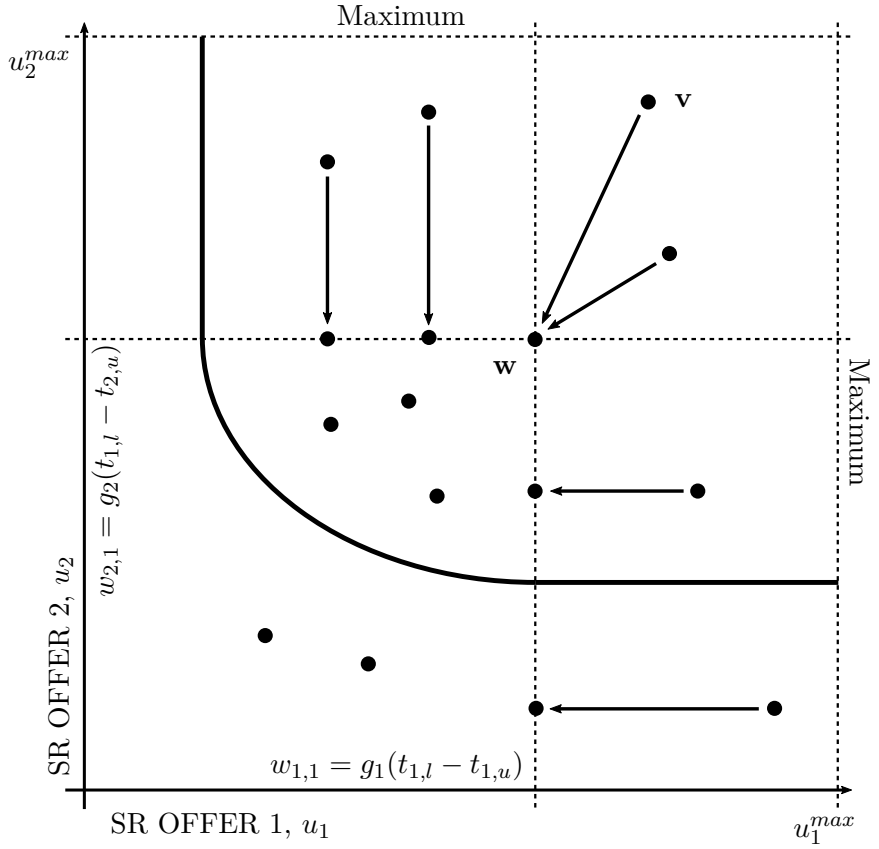


Figure 5.5 The transformation of points **v** to **w** in the example of the two SR offer problem. The arrows show the direction from **v** to **w**.

The frequency constraint Eq. 5.8 is rearranged as a function of **w**:

$$F_j(\mathbf{w}) = -2Hf_j - Rt_{j,l} + \sum_i q_i(t_{j,l} - t_{i,p}) + \sum_i (w_i(t_{j,l} - t_{i,u}) - \frac{w_i^2}{2g_i}) \tag{5.26}$$

It is a matter of following the transformations, Eq. 5.24 and 5.25, and definitions to show that $F_j(\mathbf{w}) \geq 0$ is equivalent to Eq. 5.8. To demonstrate convexity, it is to prove that $F_j(T_j((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2)) \geq 0$ given $F_j(T_j(\mathbf{v}_1)) \geq 0$ and $F_j(T_j(\mathbf{v}_2)) \geq 0$. However it

is simpler to show:

$$F_j(T_j((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2)) - (1-\lambda)F_j(T_j(\mathbf{v}_1)) - \lambda F_j(T_j(\mathbf{v}_2)) \geq 0 \tag{5.27}$$

as each term starting with inertia, risk, and then each offer can be collected individually and shown to be greater than or equal to zero. Eq. 5.27 implies $F_j(T_j((1-\lambda)\mathbf{v}_1+\lambda\mathbf{v}_2)) \geq 0$ because:

$$F_j(T_j((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2)) \geq (1-\lambda)F_j(T_j(\mathbf{v}_1)) + \lambda F_j(T_j(\mathbf{v}_2)) \geq 0$$

With a processing of evaluating each term individually, the inertia terms of Eq. 5.27 are collected together to give:

$$-2((1-\lambda)H_1 + \lambda H_2)f_j + 2(1-\lambda)H_1 f_j + 2\lambda H_2 f_j = 0 \tag{5.28}$$

and similarly for risk and IL offers:

$$-((1-\lambda)R_1 + \lambda R_2)t_{j,l} + (1-\lambda)R_1 t_{j,l} + \lambda R_2 t_{j,l} = 0 \tag{5.29}$$

$$((1-\lambda)p_{i,1} + \lambda p_{i,2})(t_{j,l} - t_{i,p}) - (1-\lambda)p_{i,1}(t_{j,l} - t_{i,p}) - \lambda p_{i,2}(t_{j,l} - t_{i,p}) = 0 \tag{5.30}$$

when $t_{i,p} \leq t_{j,l}$, otherwise all IL terms are zero and the problem is trivial. For each spinning reserve offer there are five different situations that could happen for $u_{i,1}$ and $u_{i,2}$ (the second index of u refers to the index of $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively as opposed to its reference in the two SR offer example of Figure 5.4 and accompanying text):

$$\begin{array}{lllll} 1. & u_{i,1} < w_{i,j} & \text{and} & u_{i,2} < w_{i,j} \\ 2. & u_{i,1} \geq w_{i,j} & \text{and} & u_{i,2} < w_{i,j} \\ 3. & u_{i,1} < w_{i,j} & \text{and} & u_{i,2} \geq w_{i,j} \\ 4. & u_{i,1} \geq w_{i,j} & \text{and} & u_{i,2} \geq w_{i,j} \\ 5. & t_{i,u} > t_{j,l} \end{array} \tag{5.31}$$

if $u_{i,1} = w_{i,j}$ or $u_{i,2} = w_{i,j}$ then it does not exactly matter which situation it falls into, e.g. if $u_{i,1} = w_{i,j}$ and $u_{i,2} < w_{i,j}$, then the situation could be 1 or 2. For situation 1 it is noticed that $(1-\lambda)u_{i,1} + \lambda u_{i,2} < w_{i,j}$ and similarly for situation 4, $(1-\lambda)u_{i,1} + \lambda u_{i,2} \geq w_{i,j}$, both for all $\lambda \in [0,1]$. For both situation 1 and 4, u_i does not change what side of $w_{i,j}$ it is on. Therefore collecting all the $u_{i,1}$ and $u_{i,2}$ terms for an individual i out of Eq. 5.27, and assuming situation 1, convexity is shown for this situation by the following lines of working:

$$\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)(t_{j,l}-t_{i,u})-\frac{\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2}{2g_i}$$

$$-(1-\lambda)\Big(u_{i,1}(t_{j,l}-t_{i,u})-\frac{u_{i,1}^2}{2g_i}\Big)-\lambda\Big(u_{i,2}(t_{j,l}-t_{i,u})-\frac{u_{i,2}^2}{2g_i}\Big)$$

$$=\frac{1}{2g_i}\Big((1-\lambda)u_{i,1}^2+\lambda u_{i,2}^2-\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2\Big)$$

$$=\frac{1}{2g_i}\Big(\Big((1-\lambda)-(1-\lambda)^2\Big)u_{i,1}^2-2\lambda(1-\lambda)u_{i,1}u_{i,2}+(\lambda-\lambda^2)u_{i,2}^2\Big)$$

$$=\lambda(1-\lambda)\frac{(u_{i,1}-u_{i,2})^2}{2g_i}\geq 0 \tag{5.32}$$

For situation 4:

$$w_{i,j}(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2}{2g_i}-(1-\lambda)\Big(w_{i,j}(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2}{2g_i}\Big)$$

$$-\lambda\Big(w_{i,j}(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2}{2g_i}\Big)=0 \tag{5.33}$$

For situations 2 and 3 there is a transition between the two states at λ_i, where $w_{i,j}=(1-\lambda_i)u_{i,1}+\lambda_i u_{i,2}$. For situation 2 when $\lambda\in[0,\lambda_i]$, $(1-\lambda)u_{i,1}+\lambda u_{i,2}\geq w_{i,j}$, and then when $\lambda\in[\lambda_i,1]$, $(1-\lambda)u_{i,1}+\lambda u_{i,2}\leq w_{i,j}$. To confirm Eq. 5.27, each state is evaluated separately. Therefore starting with situation 2 when $\lambda\in[0,\lambda_i]$:

$$w_{i,j}(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2}{2g_i}-(1-\lambda)\Big(w_{i,j}(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2}{2g_i}\Big)-\lambda\Big(u_{i,2}(t_{j,l}-t_{i,u})-\frac{u_{i,2}^2}{2g_i}\Big)$$

$$=\lambda\Big((w_{i,j}-u_{i,2})(t_{j,l}-t_{i,u})-\frac{w_{i,j}^2-u_{i,2}^2}{2g_i}\Big) \tag{5.34}$$

substituting in $(t_{j,l}-t_{i,u})=w_{i,j}/g_i$:

$$=\lambda\frac{2w_{i,j}(w_{i,j}-u_{i,2})-w_{i,j}^2+u_{i,2}^2}{2g_i}=\lambda\frac{(w_{i,j}-u_{i,2})^2}{2g_i}\geq 0 \tag{5.35}$$

For situation 2 when $\lambda \in [\lambda_i, 1]$:

$$\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)(t_{j,l}-t_{i,u}) - \frac{\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2}{2g_i} - (1-\lambda)\Big(w_{i,j}(t_{j,l}-t_{i,u}) - \frac{w_{i,j}^2}{2g_i}\Big)$$

$$-\lambda\Big(u_{i,2}(t_{j,l}-t_{i,u}) - \frac{u_{i,2}^2}{2g_i}\Big)$$

substituting in $(t_{j,l}-t_{i,u}) = w_{i,j}/g_i$:

$$= \frac{1}{2g_i}\Bigg(2w_{i,j}\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big) - \Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2 - (1-\lambda)w_{i,j}^2 - \lambda(2w_{i,j}u_{i,2}-u_{i,2}^2)\Bigg)$$

$$= \frac{1}{2g_i}\Bigg(2(1-\lambda)w_{i,j}u_{i,1} - \Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2 - (1-\lambda)w_{i,j}^2 + \lambda u_{i,2}^2\Bigg)$$

substituting in $w_{i,j} = (1-\lambda_i)u_{i,1} + \lambda_i u_{i,2}$:

$$= \frac{1}{2g_i}\Bigg(u_{i,1}^2\Big(2(1-\lambda)(1-\lambda_i) - (1-\lambda)^2 - (1-\lambda)(1-\lambda_i)^2\Big)$$

$$+2u_{i,1}u_{i,2}\Big((1-\lambda)\lambda_i - \lambda(1-\lambda) - (1-\lambda)\lambda_i(1-\lambda_i)\Big)$$

$$+u_{i,2}^2\Big(-\lambda^2 - (1-\lambda)\lambda_i^2 + \lambda\Big)\Bigg)$$

$$= (1-\lambda)(\lambda-\lambda_i^2)\frac{(u_{i,1}-u_{i,2})^2}{2g_i} \geq 0 \tag{5.36}$$

It is greater than or equal to zero because $\lambda - \lambda_i^2 \geq 0$. For situation 3 when $\lambda \in [0, \lambda_i]$:

$$\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)(t_{j,l}-t_{i,u}) - \frac{\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2}{2g_i} - (1-\lambda)\Big(u_{i,1}(t_{j,l}-t_{i,u}) - \frac{u_{i,1}^2}{2g_i}\Big)$$

$$-\lambda\Big(w_{i,j}(t_{j,l}-t_{i,u}) - \frac{w_{i,j}^2}{2g_i}\Big)$$

substituting in $(t_{j,l}-t_{i,u}) = w_{i,j}/g_i$:

$$= \frac{1}{2g_i}\Bigg(2w_{i,j}\Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big) - \Big((1-\lambda)u_{i,1}+\lambda u_{i,2}\Big)^2 - (1-\lambda)(2w_{i,j}u_{i,1}-u_{i,1}^2) - \lambda w_{i,j}^2\Bigg)$$

$$= \frac{1}{2g_i}\Big(2\lambda w_{i,j}u_{i,2} - \Big((1-\lambda)u_{i,1} + \lambda u_{i,2}\Big)^2 + (1-\lambda)u_{i,1}^2 - \lambda w_{i,j}^2\Big)$$

substituting in $w_{i,j} = (1-\lambda_i)u_{i,1} + \lambda_i u_{i,2}$:

$$= \frac{1}{2g_i}\Big(u_{i,1}^2\Big(-(1-\lambda)^2 + (1-\lambda) - \lambda(1-\lambda_i)^2\Big)$$
$$+2u_{i,1}u_{i,2}\Big(\lambda(1-\lambda_i) - \lambda(1-\lambda) - \lambda\lambda_i(1-\lambda_i)\Big)$$
$$+u_{i,2}^2\Big(2\lambda\lambda_i - \lambda^2 - \lambda\lambda_i^2\Big)\Big)$$

$$= \lambda\Big((1-\lambda) - (1-\lambda_i)^2\Big)\frac{(u_{i,1}-u_{i,2})^2}{2g_i} \geq 0 \tag{5.37}$$

For situation 3 when $\lambda \in [\lambda_i, 1]$:

$$w_{i,j}(t_{j,l} - t_{i,u}) - \frac{w_{i,j}^2}{2g_i} - (1-\lambda)\Big(u_{i,1}(t_{j,l} - t_{i,u}) - \frac{u_{i,1}^2}{2g_i}\Big) - \lambda\Big(w_{i,j}(t_{j,l} - t_{i,u}) - \frac{w_{i,j}^2}{2g_i}\Big)$$

$$= (1-\lambda)\Big((w_{i,j} - u_{i,1})(t_{j,l} - t_{i,u}) - \frac{w_{i,j}^2 - u_{i,1}^2}{2g_i}\Big)$$

substituting in $(t_{j,l} - t_{i,u}) = w_{i,j}/g_i$:

$$= (1-\lambda)\frac{2w_{i,j}(w_{i,j} - u_{i,1}) - w_{i,j}^2 + u_{i,1}^2}{2g_i} = (1-\lambda)\frac{(w_{i,j} - u_{i,1})^2}{2g_i} \geq 0 \tag{5.38}$$

Finally for situation 5 it is the trivial case, and so it has been shown that the space bounded by a single frequency limit is convex. Now it is left to show that the intersection of all these spaces proved so far is also convex. Define the set $S_0 \subset \mathbb{R}^N$, where $N = 2 + Np + Nu$, i.e. one variable each for inertia and risk, Np variables for IL, and Nu for SR. S_0 is the set of points that satisfy the box constraints of Eq. 5.13 and 5.14, and also the reserve requirement constraint of Eq. 5.7. For each frequency limit, define the set $S_j \subset \mathbb{R}^N$ in correspondence time $t_{j,l}$, and frequency f_j. The intersection of all these sets is expressed as:

$$S = \bigcap_{j=0}^{Nc} S_j \tag{5.39}$$

where Nc are the number of $(t_{j,l}, f_j)$ pairs after time zero. It is possible for there

to be no pairs and just have a single frequency limit at time zero, but this is the trivial case, so assume there is at least one pair. It has to be shown $\forall \mathbf{v}_1, \mathbf{v}_2 \in S$ that $\forall \lambda \in [0,1] \quad \mathbf{v} = (1-\lambda)\mathbf{v}_1 + \lambda \mathbf{v}_2 \in S$. Since each vector, $\mathbf{v}_1$ and $\mathbf{v}_2$, is in S then each vector is in each S_j. Therefore since each S_j is convex then $\mathbf{v} \in S_j$, and because this is true for every S_j it must also be that $\mathbf{v} \in S$. Therefore S is convex.

5.2.3 Minimum Frequency Constraint

This section proves that the subset $S_F \subseteq S$ is convex where Eq. 5.9, 5.10, and 5.11 are satisfied as well as all the other constraints. This proof requires several new definitions and steps.

The First Approach

The first and unsuccessful approach that was tried in obtaining a convex space with Eq. 5.11 was to include t_{min} as a new component to the vector $\mathbf{v}$, so that $\mathbf{v} = [H, R, t_{min}, p_1, \ldots, p_{Np}, u_1, \ldots, u_{Nu}]$. The definition of t_{min} in Eq. 5.9 and 5.10 has to be incorporated through additional constraints on the feasible space. A constraint where the amount of reserve has to be equal to the risk at time t_{min} has to be included to define t_{min}. This constraint comes in two forms. The first form is when t_{min} happens to equal at least one $t_{i,p}$, and requires two equations:

$$-R + \sum_{t_{i,p}<t_{min}} p_i + \sum_{t_{i,u}\leq t_{min}} U_i(t_{min}; u_i) \leq 0 \tag{5.40}$$

$$-R + \sum_{t_{i,p}\leq t_{min}} p_i + \sum_{t_{i,u}\leq t_{min}} U_i(t_{min}; u_i) \geq 0 \tag{5.41}$$

the difference between these two equations is the addition of IL at time t_{min} to cover the risk. The second form of the constraint is when there is no $t_{i,p} = t_{min}$, and requires only one equation:

$$-R + \sum_{t_{i,p}<t_{min}} p_i + \sum_{t_{i,u}\leq t_{min}} U_i(t_{min}; u_i) = 0 \tag{5.42}$$

It can be proven that a space defined by the intersection of S and Eq. 5.11 without Eq. 5.40-5.42 is not convex. Further restricting the space with Eq. 5.40-5.42 to obtain the feasible solution space is not likely to be convex either. The main reason why convexity is lost is because of the quadratic forms Rt_{min}, $t_{min}p_i$ and $t_{min}U_i(t_{min}; u_i)$ of Eq. 5.11. The approach of including t_{min} as a component of $\mathbf{v}$ does not work. Therefore t_{min} has to be removed from Eq. 5.11, and $U_i(t_{min}; u_i)$ has to be defined by a variable.

Minimum Frequency Constraint Transformation

To remove t_{min} from the right hand side of Eq. 5.11, a similar transformation is applied to $\mathbf{v}$ as it was in Eq. 5.23. The similar principle that is applied is that reserve that responds after t_{min} has no influence on whether the minimum frequency constraint is satisfied. This transformation is expressed as:

$$\mathbf{w} = T_m(\mathbf{v}) \tag{5.43}$$

the subscript m does not reference a series of transformations as j did for each frequency limit in Eq. 5.23, but is specific to the single minimum frequency constraint. The transformation rules are also very similar, the only difference is $t_{j,l}$ is replaced by t_{min}:

$$w_i = \begin{cases} u_i & t_{i,u} + u_i/g_i < t_{min} \\ w_{i,m} & t_{i,u} + u_i/g_i \geq t_{min} \quad \text{and} \quad t_{i,u} \leq t_{min} \\ 0 & t_{i,u} > t_{min} \end{cases} \tag{5.44}$$

where $w_{i,m} = g_i(t_{min} - t_{i,u})$. The main difference comes from handling IL when there exists $t_{i,p} = t_{min}$:

$$q_i = \begin{cases} p_i & t_{i,p} < t_{min} \\ \rho_i & t_{i,p} = t_{min} \\ 0 & t_{i,p} > t_{min} \end{cases} \tag{5.45}$$

where ρ_i solves:

$$-R + \sum_{t_{i,p}=t_{min}} \rho_i + \sum_{t_{i,p}<t_{min}} p_i + \sum_{t_{i,u}\leq t_{min}} U_i(t_{min}; u_i) = 0 \tag{5.46}$$

It appears that the value of ρ_i is unimportant in the satisfying of Eq. 5.11, as $t_{min} - t_{i,p} = 0$ in Eq. 5.11, and $\rho_i = 0$ in Eq. 5.45, except it is useful to set it in Eq. 5.46 for purposes of removing t_{min} from Eq. 5.11. The transformation in its current form does not always uniquely define $\mathbf{w}$, when there is more than one IL offer with $t_{i,p} = t_{min}$. To remove this issue, each IL offer with the same time $t_{i,p}$ can be combined into one offer with a total maximum. This is possible because the IL term of Eq. 5.11 has the property that:

$$\sum_{t_{i,p}\,\text{unique}} p_i(t_{min} - t_{i,p}) = (t_{min} - t_{i,p}) \sum_{t_{i,p}\,\text{unique}} p_i \tag{5.47}$$

i.e. IL offers with the same $t_{i,p}$ are indistinguishable with respect to satisfying Eq. 5.11. Therefore they can be combined into one equivalent offer. The minimum frequency

constraint of Eq. 5.11 is transformed into the following form:

$$-Rt_{min} + \sum_i q_i(t_{min} - t_{i,p}) + \sum_i (w_i(t_{min} - t_{i,u}) - \frac{w_i^2}{2g_i}) \geq 2Hf_{lim}(t_{min}) \tag{5.48}$$

but it is also known that through Eq. 5.9, 5.10, and 5.46:

$$-R + \sum_i q_i + \sum_i w_i = 0 \tag{5.49}$$

and now t_{min} can be removed from the minimum frequency constraint, and multiplying both sides of Eq. 5.48 by -1, the function $A(\mathbf{w})$ is defined:

$$A(\mathbf{w}) = \sum_i q_i t_{i,p} + \sum_i (w_i t_{i,u} + \frac{w_i^2}{2g_i}) \leq -2Hf_{lim}(t_{min}) \tag{5.50}$$

For example, consider the problem of two SR offers, but now excluding the step change in frequency limit, and just having the initial frequency limit f_0. The transformation T_m is shown in Figure 5.6. The different regions in the problem space are separated by an important point, the dispatch along the reserve requirement line at which the earliest time the minimum frequency can occur ($t_{min,0}$). Any point along the reserve requirement line, away from this $t_{min,0}$ point, the minimum time increases. This point is found by solving for $t_{min,0}$ in:

$$g_1(t_{min,0} - t_{1,u}) + g_2(t_{min,0} - t_{2,u}) = R$$

$$t_{min,0} = \frac{R + g_1 t_{1,u} + g_2 t_{2,u}}{g_1 + g_2} \tag{5.51}$$

where $u_1 = g_1(t_{min,0} - t_{1,u})$ and $u_2 = g_2(t_{min,0} - t_{2,u})$ to derive the first equation $u_1 + u_2 = R$. When moving away from this point, along the reserve requirement line, there is always one SR offer decreasing. This requires the other to increase, and requires more time for that offer to ramp up to that output. Hence the minimum time is increasing in either direction along the reserve requirement line away from the $t_{min,0}$ point. Above the reserve requirement line the transformation brings $\mathbf{v}$ back onto the line, as required by Eq. 5.44 and 5.45, so that the reserve dispatched before and up to t_{min} is the key to determining whether the minimum frequency constraint is satisfied.

The feasible region for this example is found by rearranging Eq. 5.50.

$$w_1 t_{1,u} + \frac{w_1^2}{2g_1} + w_2 t_{2,u} + \frac{w_2^2}{2g_2} \leq -2Hf_0$$

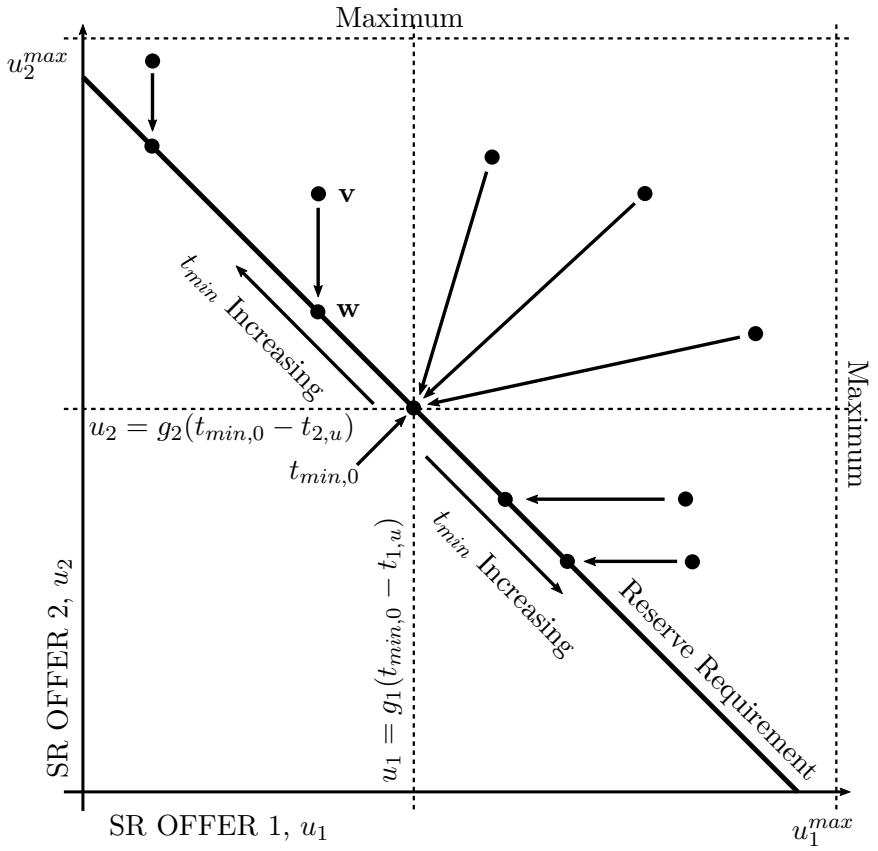


Figure 5.6 The transformation of the problem space to assess if the minimum frequency constraint is satisfied. For the two SR offer example.

$$\frac{(w_1 + g_1 t_{1,u})^2}{2g_1} + \frac{(w_2 + g_2 t_{2,u})^2}{2g_2} \leq -2Hf_0 + \frac{1}{2}g_1 t_{1,u}^2 + \frac{1}{2}g_2 t_{2,u}^2 \tag{5.52}$$

which is an equation for an ellipse, but centred at $(-g_1 t_{1,u}, -g_2 t_{2,u})$ outside of the box. Only a small proportion of the ellipse is inside the reserve bounds, as shown by the red curve in Figure 5.7. The feasible region is not all the points inside the ellipse, rather only the ones on the reserve requirement line. Then all the points that transform to that section by T_m are also part of the feasible space, as shown in Figure 5.7 by the shaded region. This example is convex and can be solved as a single LP problem, but most practical problems are more complicated than this.

Generally, to prove convexity is to show for any two $\mathbf{v}_1$, $\mathbf{v}_2$ that satisfy the minimum frequency constraint, i.e. given $A(T_m(\mathbf{v}_1)) \leq -2H_1 f_{lim}(t_{min}(\mathbf{v}_1))$ and $A(T_m(\mathbf{v}_2)) \leq -2H_2 f_{lim}(t_{min}(\mathbf{v}_2))$, that $\forall \lambda \in [0,1]$:

$$A\Big(T_m\Big((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2\Big)\Big) \leq -2\Big((1-\lambda)H_1 + \lambda H_2\Big) f_{lim}\Big(t_{min}\Big((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2\Big)\Big) \tag{5.53}$$

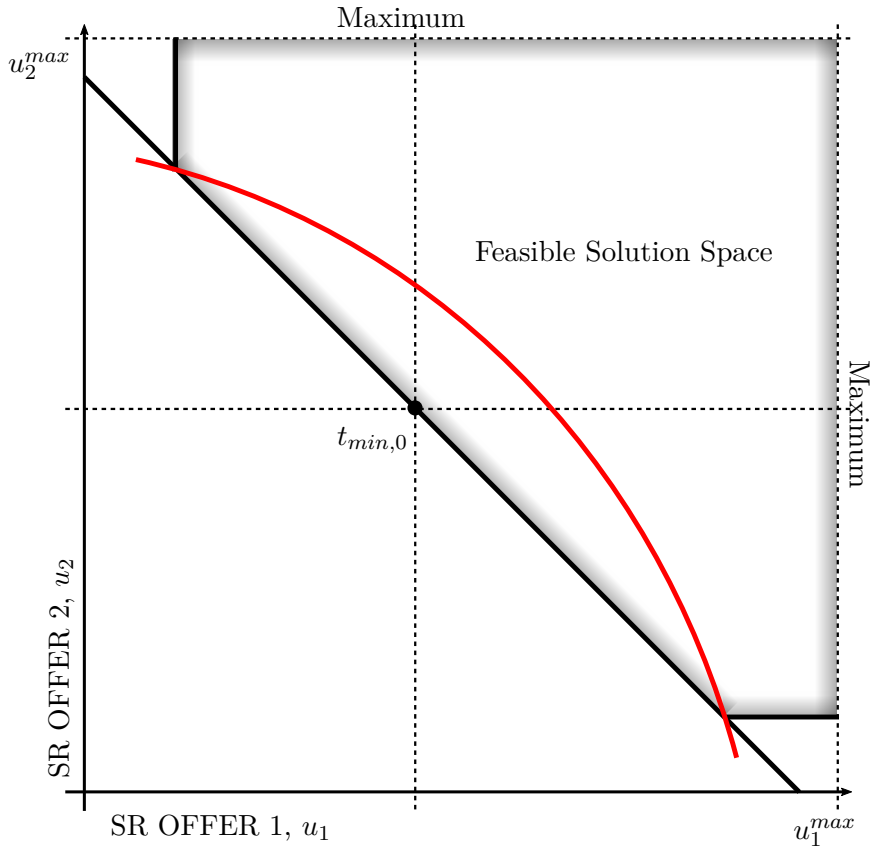


Figure 5.7 The feasible region for a simple two SR offer problem. The red line is the minimum frequency constraint determining the feasible points on the reserve requirement line.

Simplifying the above equation as $A(\lambda) \leq -2H(\lambda)f(\lambda)$, the method of proving convexity is by demonstrating $A(\lambda)$, $H(\lambda)$, and $f(\lambda)$ have specific shapes so that the inequality is always met.

The Time of the Minimum Frequency

This section describes the nature of $t_{min}(\lambda)$ as λ varies from 0 to 1. The full description is provided in Appendix A, only an abstraction is given here. The time of minimum frequency is when the total reserve is equal to the risk. If the total reserve before t_{min} increases with λ, then minimum time is brought forward. If the total reserve before t_{min} reduces then the minimum time delays. To mathematically describe t_{min}, the domain of λ, $[0, 1]$, is separated into a finite number of regions by transition markers λ_i. For $\lambda \in (\lambda_i, \lambda_{i+1})$, $t_{min}(\lambda)$ is linear. At λ_i the gradient of $t_{min}(\lambda)$ can change, or there is a jump, or both. The jumps occur when the frequency transient, $f(t)$, is flat at the minimum for a period of time after t_{min}. This is when $df/dt = 0$ for longer than an instance, and only occurs when no SR is ramping after t_{min} for a finite duration.

The minimum time does not have any random assortment of linear segments, but obeys a specific rule: if t_{min} has started to increase, then it can no longer decrease, which is more accurately said in Result 1, and proven in Appendix A:

Result 1: *For all* $\lambda_1, \lambda_2 \in [0, 1)$ *where* $0 \leq \lambda_1 < \lambda_2 < 1$, *if* $t_{min}(\lambda_1) < t_{min}(\lambda_2)$ *then for all* $\alpha, \beta \in [\lambda_2, 1]$ *where* $\lambda_2 \leq \alpha \leq \beta \leq 1$, $t_{min}(\alpha) \leq t_{min}(\beta)$. *That is when* t_{min} *starts to increase it will always be monotonically increasing after that.*

The general profile of t_{min} is presented in Figure 5.8. In this example, a minimum is reached, but this is not necessary, as it can be either decreasing for the whole of λ or increasing. Between λ_1 and λ_2, an IL offer is the critical offer that satisfies the reserve requirement, similarly at λ_7 and λ_8. At λ_3 there is a jump because there is no SR offer ramping after t_{min} and for this value of λ_i.

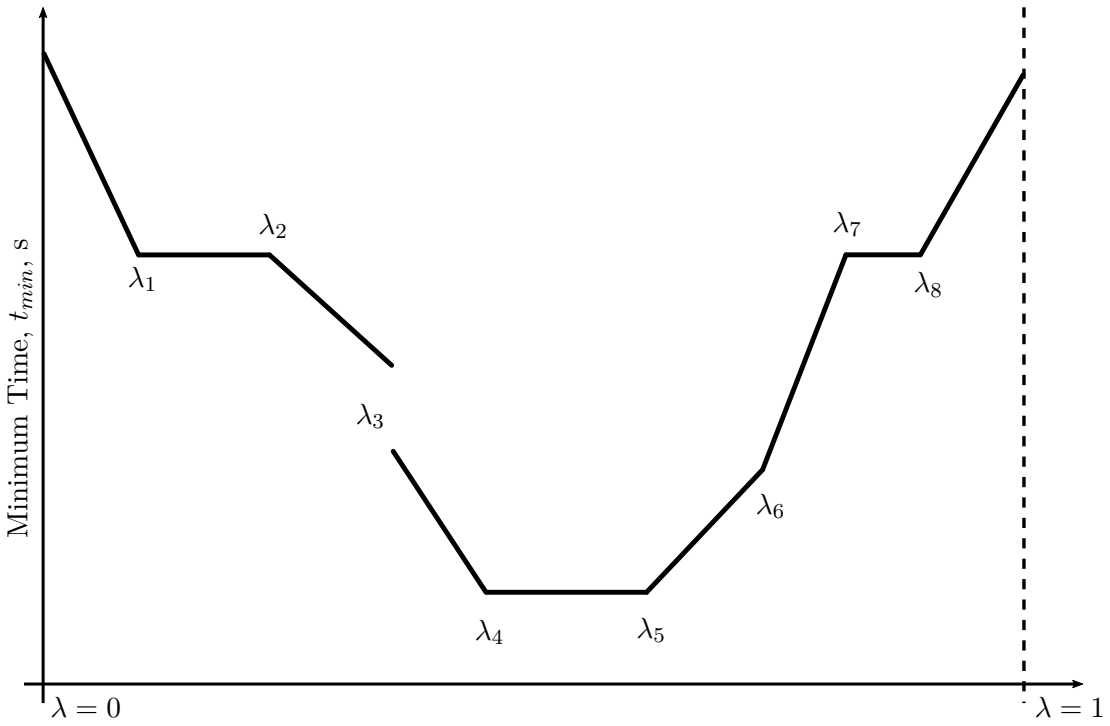


Figure 5.8 The general profile of t_{min} with λ

The full properties of Result 1 are not necessary in simplify Eq. 5.53, but provides some explanation as to how f_{lim} varies with λ. The frequency limit can decrease several times in a sequence, but once it has increased it can no longer decrease again. To simplify Eq. 5.53, define the points ζ_k in the domain of λ where $t_{min}(\zeta_k) = t_{j,l}$. If for one transition, λ_i, there is a jump and $t_{min}(\lambda_i^-) \leq t_{j,l} \leq t_{min}(\lambda_i^+)$ or $t_{min}(\lambda_i^-) \geq t_{j,l} \geq$

$t_{min}(\lambda_i^+)$ then include λ_i with ζ_k. This is seen in Figure 5.9 for λ_3, where $t_{2,l}$ intersects the discontinuity, and results in the creation of ζ_2. If there is a whole continuum of $t_{min}(\lambda) = t_{j,l}$ then only include the first and last values of the continuum with ζ_k. This is seen in Figure 5.9 where between λ_4 and λ_5, $t_{min}(\lambda) = t_{1,l}$, hence only two points ζ_3 and ζ_4 are required. Therefore there is a finite number of ζ_k.

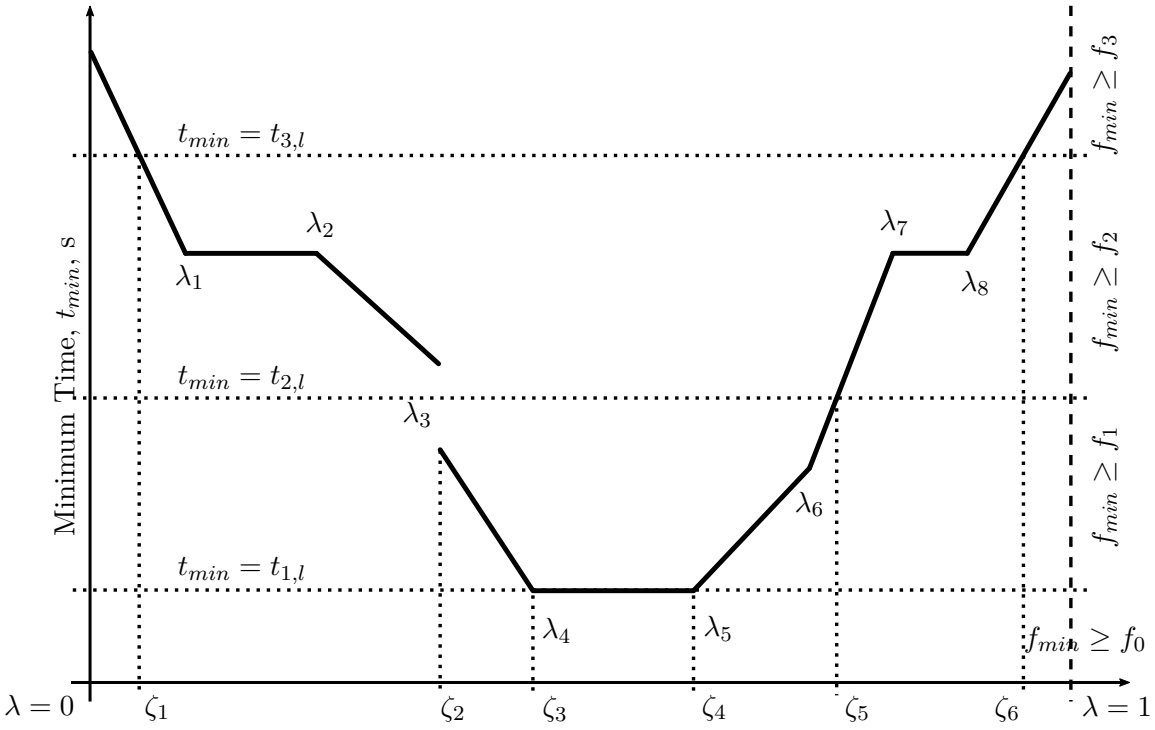


Figure 5.9 The possible locations of ζ_k in the domain of λ for one possible set of times for the frequency limit step changes, where $f_{min} = f(t_{min})$.

For each time $\lambda = \zeta_k$ it is known that Eq. 5.53 is satisfied:

$$A\Big(T_m\Big((1-\zeta_k)\mathbf{v}_1 + \zeta_k\mathbf{v}_2\Big)\Big) \leq -2\Big((1-\zeta_k)H_1 + \zeta_k H_2\Big) f_{lim}\Big(t_{min}\Big((1-\zeta_k)\mathbf{v}_1 + \zeta_k\mathbf{v}_2\Big)\Big) \tag{5.54}$$

because this constraint is equivalent to one of the frequency limit constraints, Eq. 5.8, for a particular $t_{j,l}$ and f_j. Showing the convexity of the minimum frequency constraint is simplified to show for all k, and for all $\lambda \in (\zeta_k, \zeta_{k+1})$ that:

$$A\Big(T_m\Big((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2\Big)\Big) \leq -2\Big((1-\lambda)H_1 + \lambda H_2\Big) f_k \tag{5.55}$$

including the domains $(0, \zeta_1)$ and $(\zeta_{Nk}, 1)$, where Nk is the last index of ζ_k. The frequency limit f_k is found by picking a λ where $\zeta_k < \lambda < \zeta_{k+1}$: $f_k = f_{lim}(t_{min}(\lambda))$.

While knowing through Eq. 5.54 that:

$$A\Big(T_m\Big((1-\zeta_k)\mathbf{v}_1+\zeta_k\mathbf{v}_2\Big)\Big) \leq -2\Big((1-\zeta_k)H_1+\zeta_k H_2\Big)f_k \tag{5.56}$$

$$A\Big(T_m\Big((1-\zeta_{k+1})\mathbf{v}_1+\zeta_{k+1}\mathbf{v}_2\Big)\Big) \leq -2\Big((1-\zeta_{k+1})H_1+\zeta_{k+1} H_2\Big)f_k \tag{5.57}$$

Therefore the right hand side of Eq. 5.53 is simplified to a linear segment, as seen by the right hand side of Eq. 5.55 being a linear function of λ.

Area Function

The function $A(\lambda)$, as already defined in Eq. 5.50, is called the area function because it is the result of finding the area between the power imbalance curve and the axis, as it is shown in Figure 5.10. The power imbalance curve is the right hand side of Eq. 5.4. The units of $A(\lambda)$ are in MW s.

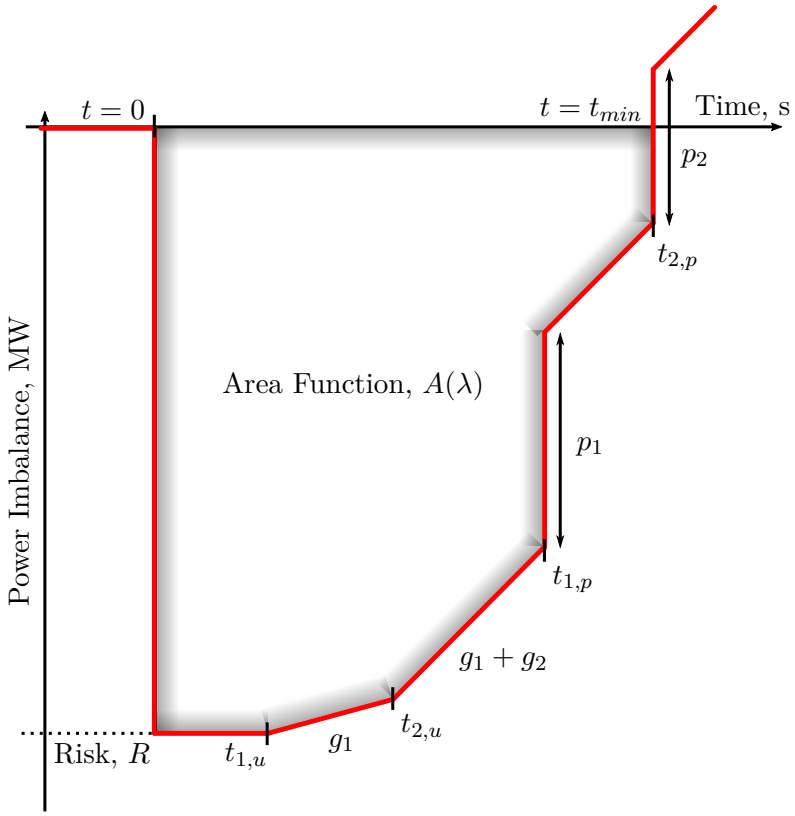


Figure 5.10 The area function, $A(\lambda)$, highlighted as the area below the axis. Four reserve response are shown for this example: two SR offers and two IL offers. Note that $A(\lambda)$ is the absolute area.

In Appendix A, it is shown that $A(\lambda)$ is continuous, and has the important property:

Result 2: *The second derivative of $A(\lambda)$ is a non-negative constant, $d^2A/d\lambda^2 = \alpha_i \geq 0$ for λ in an open interval $(\lambda_{i-1}, \lambda_i)$, and for $\lambda = \lambda_i$:*

$$\left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_i^-} \leq \left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_i^+} \tag{5.58}$$

It will be shown from this result that Eq. 5.55 is always satisfied, and the feasible solution space is convex. Firstly express the left hand side of Eq. 5.56 and 5.57 as A_1 and A_2 respectively, and the right hand side as F_1 and F_2, so that $A_1 \leq F_1$ and $A_2 \leq F_2$. Draw a line between A_1 and A_2, and a line between F_1 and F_2 with the parameter δ:

$$\lambda = (1-\delta)\zeta_k + \delta\zeta_{k+1}$$

$$A_{line}(\delta) = (1-\delta)A_1 + \delta A_2 \tag{5.59}$$

$$F_{line}(\delta) = (1-\delta)F_1 + \delta F_2 \tag{5.60}$$

The proof is by contradiction, assume that there exists a point, δ^*, where $A(\lambda(\delta^*)) > A_{line}(\delta^*)$, and with the potential for $A(\lambda(\delta^*)) > (1-\delta^*)F_1 + \delta^* F_2$. By assuming this, it is shown that it is impossible for $A(\lambda)$ to both satisfy Result 2 and $A(\zeta_{k+1}) = A_2$, therefore $A(\lambda((\delta))) \leq A_{line}(\delta)) \leq F_{line}(\delta))$ and the problem space is convex. To show this impossibility, start by shortening $A(\lambda(\delta^*))$ to A^*. These points, δ^* and A^*, are plotted in Figure 5.11, along with the lines A_{line} and F_{line}.

The minimum possible gradient of $A(\lambda)$ at δ^* while retaining the properties of result 2 is:

$$\min\{\frac{dA(\delta^*)}{d\lambda}\} = \frac{A^* - A_1}{\lambda^* - \zeta_k} \tag{5.61}$$

if $(1-\delta^*)\zeta_k + \delta^*\zeta_{k+1} = \lambda_i$ a transition point, then the minimum gradient is the one approaching from the left. To show this is the minimum possible gradient, an optimisation problem is formulated. The objective is to minimise the derivative:

$$\frac{dA(\delta^*)}{d\lambda} = \frac{dA(\zeta_k)}{d\lambda} + \sum_{i=1}^{N_T}\beta_i + \sum_{i=1}^{N_T+1}\int_{\lambda_{i-1}}^{\lambda_i}\frac{d^2A}{d\lambda^2}d\lambda$$

and after evaluating the integral:

$$\frac{dA(\delta^*)}{d\lambda} = \frac{dA(\zeta_k)}{d\lambda} + \sum_{i=1}^{N_T}\beta_i + \sum_{i=1}^{N_T+1}\alpha_i(\lambda_i - \lambda_{i-1}) \tag{5.62}$$

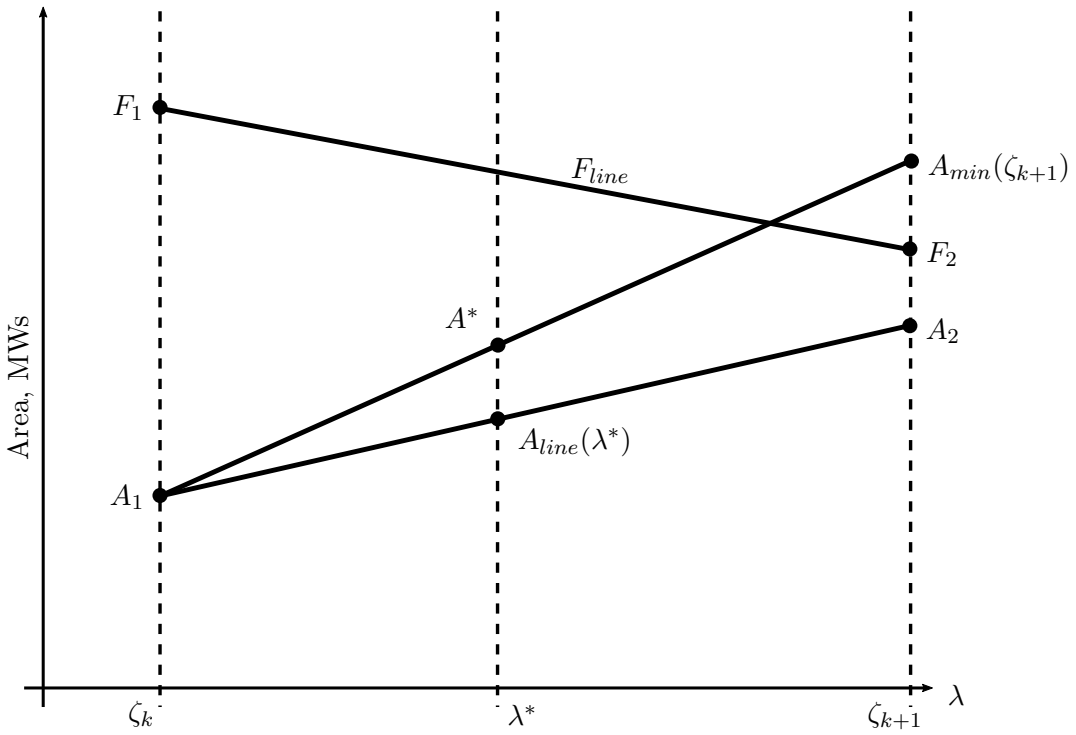


Figure 5.11 The points used to prove that $A(\lambda)$ is always below the line between A_1 and A_2, by proving A^* can never be above $A_{line}(\lambda(\delta^*))$.

where there can be any finite number of transitions, N_T. Note that just for this optimisation that the indices of λ_i are shifted so that $\lambda_0 = \zeta_k$ and $\lambda_{N_T+1} = (1-\delta^*)\zeta_k + \delta^*\zeta_{k+1}$. The acceleration term, α_i, is the same from Result 2, and β_i:

$$\beta_i = \left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_i^+} - \left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_i^-} \tag{5.63}$$

in this optimisation $dA(\zeta_k)/d\lambda$, α_i, and β_i are the variables. There is one equality constraint, the area function at δ^* has to equal A^*:

$$A^* = A(\zeta_k) + \int_{\lambda=\zeta_k}^{\lambda^*} dA(\lambda)$$

skipping the long process of evaluating this integral, the result is:

$$A^* = A(\zeta_k) + \frac{dA(\zeta_k)}{d\lambda}(\lambda^* - \zeta_k) + \sum_{i=1}^{N_T} \Big(\beta_i + \alpha_i(\lambda_i - \lambda_{i-1})\Big)(\lambda^* - \lambda_i) + \sum_{i=1}^{N_T+1} \frac{\alpha_i}{2}(\lambda_i - \lambda_{i-1})^2 \tag{5.64}$$

Result 2 gives the inequality constraints, $\alpha_i \geq 0$ and $\beta_i \geq 0$.

The desired KKT conditions for this optimisation are:

$$1 + \Lambda(\lambda^* - \zeta_k) = 0 \tag{5.65}$$

$$\forall i \leq N_T \quad 1 + \Lambda(\lambda^* - \lambda_i) - M_{\beta i} = 0 \tag{5.66}$$

$$\forall i \leq N_T \quad (\lambda_i - \lambda_{i-1}) + \Lambda\Big((\lambda_i - \lambda_{i-1})(\lambda^* - \lambda_i) + \frac{1}{2}(\lambda_i - \lambda_{i-1})^2\Big) - M_{\alpha i} = 0 \tag{5.67}$$

$$i = N_T + 1 \quad (\lambda_i - \lambda_{i-1}) + \frac{1}{2}\Lambda(\lambda_i - \lambda_{i-1})^2 - M_{\alpha i} = 0 \tag{5.68}$$

$$X_i M_{\beta i} = 0 \quad \text{and} \quad \alpha_i M_{\alpha i} = 0 \tag{5.69}$$

$$\beta_i \geq 0 \quad \alpha_i \geq 0 \quad M_{\beta i} \geq 0 \quad M_{\alpha i} \geq 0 \tag{5.70}$$

where Λ, $M_{\beta i}$, and $M_{\alpha i}$ are the KKT multipliers. The first multiplier can be found directly by rearranging Eq. 5.65:

$$\Lambda = -\frac{1}{\lambda^* - \zeta_k} \tag{5.71}$$

The second set of multipliers can be found by substituting Eq. 5.71 into Eq. 5.66 and rearranging for $M_{\beta i}$:

$$M_{\beta i} = \frac{\lambda_i - \zeta_k}{\lambda^* - \zeta_k} > 0 \tag{5.72}$$

therefore $\beta_i = 0$ so that Eq. 5.69 is held. Substituting Λ into Eq. 5.67 and 5.68:

$$\forall i \leq N_T + 1 \quad M_{\alpha i} = (\lambda_i - \lambda_{i-1})\frac{\frac{1}{2}(\lambda_i + \lambda_{i-1}) - \zeta_k}{\lambda^* - \zeta_k} > 0 \tag{5.73}$$

therefore $\alpha_i = 0$ so that Eq. 5.69 is satisfied. From Eq. 5.62, $dA(\delta^*)/d\lambda = dA(\zeta_k)/d\lambda$, and by Eq. 5.64, Eq. 5.61 is proven correct. The shape of the optimal $A(\lambda)$ is a line between the two points, $A(\zeta_k)$ and $A(\lambda^*)$ to get the lowest derivative possible at λ^*. The lowest possible value for $A(\zeta_{k+1})$ is found by extending the optimal line:

$$A_{min}(\zeta_{k+1}) = A(\zeta_k) + (\zeta_{k+1} - \zeta_k)\frac{A^* - A_1}{\lambda^* - \zeta_k} \tag{5.74}$$

The actual value for area at ζ_{k+1} is A_2:

$$A_2 = A(\zeta_k) + (\zeta_{k+1} - \zeta_k)\frac{A_{line}(\lambda^*) - A_1}{\lambda^* - \zeta_k} \tag{5.75}$$

Since $A^* > A_{line}(\lambda^*)$ it implies $A_{min}(\zeta_{k+1}) > A_2$, a contradiction as it should be possible for them to be equal. Hence it is not possible for $A^* > A_{line}(\lambda^*)$, rather $A^* \leq A_{line}(\lambda^*)$. Connecting back to the limits, F_1 and F_2, the line between these points is always greater than A_{line}. Therefore it is shown that Eq. 5.55 is always satisfied by reformulating it as follows:

$$\forall \delta \in [0,1] \quad A(\lambda(\delta)) \leq (1-\delta)A_1 + \delta A_2 \leq (1-\delta)F_1 + \delta F_2 \tag{5.76}$$

and it is proved that the feasible space is convex.

5.3 SOLVING METHODOLOGY

The optimal solution cannot be found by solving a single set of equations, fortunately the feasible solution space can be broken up into a finite number of convex regions. This section explains the methodology of moving from one region to another, solving for the local minimum in each region, in order to find the global minimum. Since the whole space is convex, as proved before, only a small fraction of the regions have to be solved, thereby making the implementation practical for market operations which have restrictions on solve times.

5.3.1 Region Formulation

The reason why the problem cannot be solved with a single set of equations is alluded to in Section 5.2.2 in trying to prove the convexity of the frequency constraints. Consider again the simple problem of optimising two SR offers, as shown in Figure 5.4. In the bottom left quadrant an ellipse defines the region, in the top left and bottom right a straight line, and in the top right there is no curve. The equations for all these lines cannot all be incorporated into a single solver because of the ellipse. The equation that defines the ellipse proceeds into the other quadrants and closes off parts of the feasible region. Therefore the optimal solution that is found by solving with all equations may not be optimal, as the optimum may be in the region outside the ellipse but still in the feasible space.

To reiterate this point, consider another example: optimizing one SR offer and one IL offer with one step change in frequency limit at $t_{1,l}$. For this example let $t_{1,u} < t_{1,p} < t_{1,l}$. There are two forms of the frequency constraint, Eq. 5.26, in terms of p_1 and u_1 (not their transformation variables q_1 and w_1):

$$-2Hf_1 - Rt_{1,l} + p_1(t_{1,l} - t_{1,p}) + u_1(t_{1,l} - t_{1,u}) - \frac{u_1^2}{2g_1} \geq 0 \quad \text{when} \quad u_1 \leq g_1(t_{1,l} - t_{1,u}) \tag{5.77}$$

$$-2Hf_1 - Rt_{1,l} + p_1(t_{1,l} - t_{1,p}) + \frac{1}{2}g_1(t_{1,l} - t_{1,u})^2 \geq 0 \quad \text{when} \quad u_1 \geq g_1(t_{1,l} - t_{1,u}) \tag{5.78}$$

The first constraint describes a region constrained by a parabola, and the second by a straight line. The feasible region is presented in Figure 5.12 by the space above solid frequency limit line. The equation for the parabola is best understood by rearranging Eq. 5.77 through completing the square:

$$p_1 \geq \frac{\Big(u_1 - g_1(t_{1,l} - t_{1,u})\Big)^2}{2g_1(t_{1,l} - t_{1,p})} + \frac{4Hf_1 + 2Rt_{l,1} - g_1(t_{1,l} - t_{1,u})^2}{2(t_{1,l} - t_{1,u})} \tag{5.79}$$

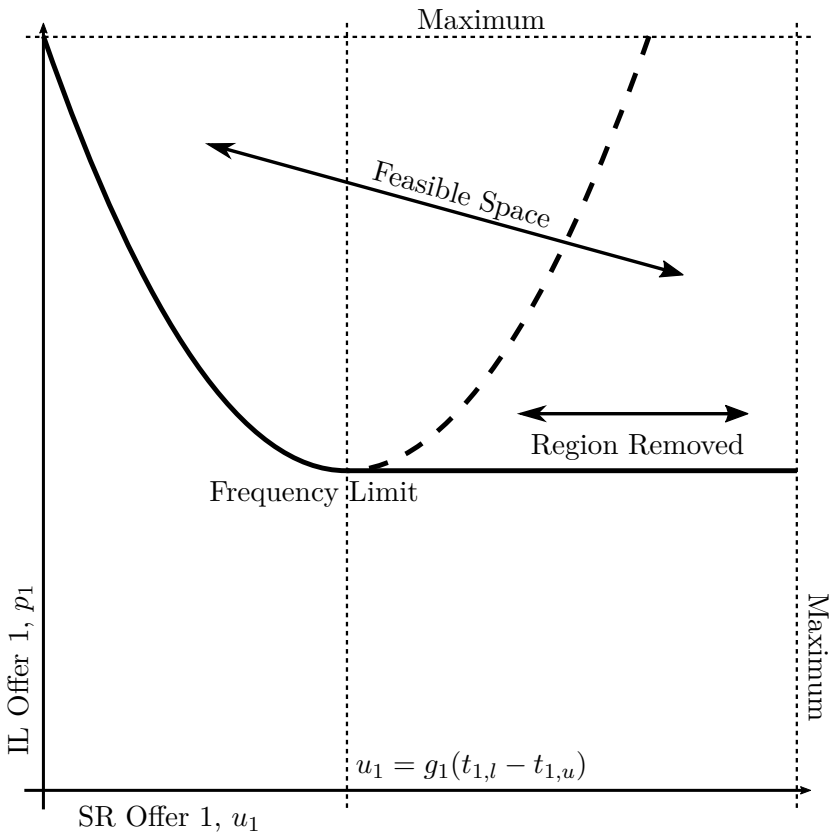


Figure 5.12 Feasible space for optimising one SR offer and one IL offer, demonstrating why it is not possible to combine all the frequency constraint equations into a single solver.

If Eq. 5.79 is allowed to be valid for $u_1 \geq g_1(t_{1,l} - t_{1,u})$, then as shown in Figure 5.12 by Region Removed, to the left of the dotted line, there is a portion of the feasible space missing. The optimal solution that can been found for this reduced space may not be the global solution in this case. Hence each region has to be solved separately.

The number of regions is determined by how many step changes in the frequency limit occur while SR is ramping up. For example consider the reserve output over time for one SR offer in Figure 5.13. There are two frequency limit step changes, $t_{2,l}$ and $t_{3,l}$ while the SR offer is ramping, so there are three regions. The first step change, $t_{1,l} \leq t_{1,u}$, and the last step change, $t_{4,l} \geq t_{1,e}^{max}$, do not contribute to making regions. If there are Nu SR offers and Nc frequency limit step changes, then the maximum number of regions is $(Nc+1)^{Nu}$. With 10 SR offers and 2 step changes there can be up to 60,000 regions to search through. If reserve and energy is co-optimised like it is for SPD in the New Zealand electricity market, each region optimisation requires roughly one second to compute, then the total time to solve the whole space would be 17 hours, clearly inappropriate for a five minute dispatch cycle. Therefore it is important to develop a solving methodology that does require to solve all these regions.

IL offers do not contribute any regions as they ramp instantly. The minimum frequency constraint, like the frequency limits, also creates regions. Not just when

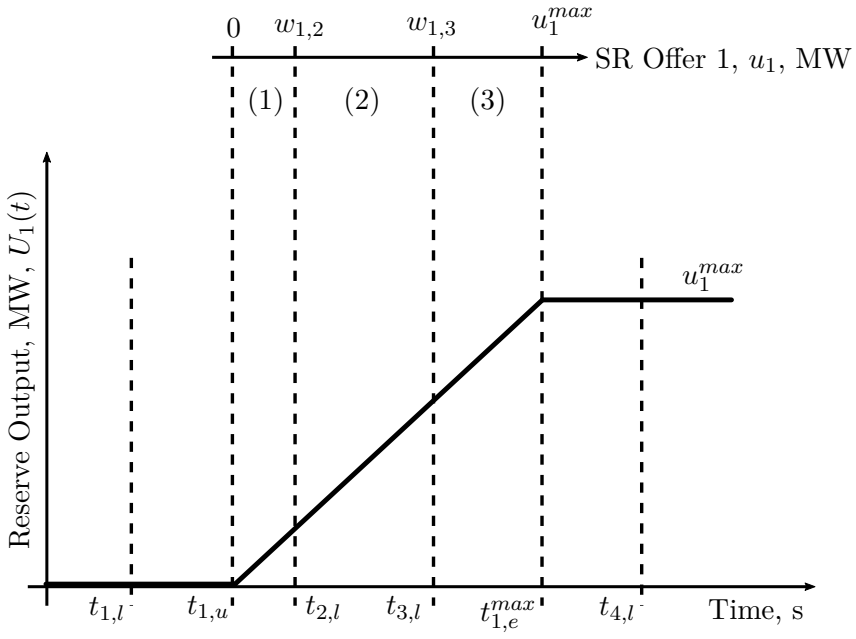


Figure 5.13 The regions a SR offer is separated into because of the step changes in the frequency limit occurring while it is ramping.

t_{min} happens to be when a SR offer is ramping, and adds another region to each SR offer, but that t_{min} varies as well. Therefore there is a region for each possible range t_{min} can vary between important times such as $t_{i,p}$, $t_{i,u}$, etc. It is possible for t_{min} to be a variable in each region and for that region to be convex. The first approach of incorporating t_{min} as a component of $\mathbf{v}$ and restricting the space by Eq. 5.11 failed because of the unwanted quadratic terms, but through Eq. 5.50 these terms are removed and it is possible to prove convexity. This is shown in the next section, the rest of this section formulates equations for each region and developing a better estimate on the total number of regions in a given problem.

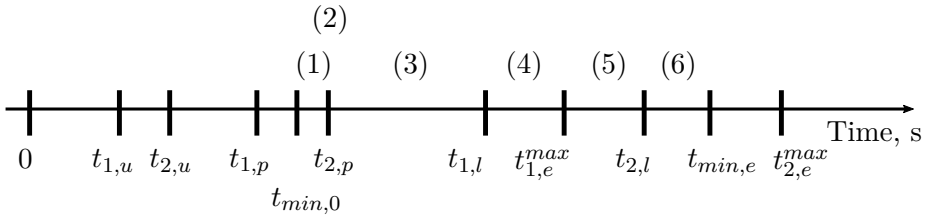


Figure 5.14 Time line of important time periods that determine each time based region.

Consider the time line in Figure 5.14, marked with each time IL is initiated, SR starts and stops ramping (the maximum possible), and each time there is a step change in the frequency limit. This is shown for an example of two SR offers, two IL offers, and two step changes in the frequency limit function in Figure 5.14. The minimum

time is to be feasible for the whole time range, $t_{min,0} \leq t_{min} \leq t_{min,e}$. For this example there are six different groups of time differentiated regions. Five of them are similar, call them the SR type, and (2) is unique, call it the IL type. (1), (3), (4), (5), and (6) are called the SR type because SR offers are the critical reserve to cover the risk. For this type t_{min} is allowed to vary over a domain. For IL type, t_{min} is held constant, as there exists $t_{i,p} = t_{min}$.

The set of equations for the IL type region are:

$$\text{minimise} \quad \sum_i c_{i,p} p_i + \sum_i c_{i,u} u_i \tag{5.80}$$

$$\text{subject to} \quad p_m + \sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i (t_{min} - t_{i,u}) = R \tag{5.81}$$

$$p_m t_{min} + \sum_{i \in Q_B} p_i t_{i,p} + \sum_{i \in W_B} (u_i t_{i,u} + \frac{u_i^2}{2g_i}) + \sum_{i \in W_T} \frac{g_i}{2} (t_{min}^2 - t_{i,u}^2) \leq -2Hf_{lim} \tag{5.82}$$

$$\forall j, \quad -Rt_{j,l} + \sum_{i \in Q_{B,j}} p_i (t_{j,l} - t_{i,p}) + \sum_{i \in W_{B,j}} \left(u_i (t_{j,l} - t_{i,u}) - \frac{u_i^2}{2g_i} \right) + \sum_{i \in W_{T,j}} \frac{g_i}{2} (t_{j,l} - t_{i,u})^2 \geq 2Hf_j \tag{5.83}$$

$$0 \leq p_m \leq \sum_{i \in Q_E} p_i \tag{5.84}$$

$$\forall i, \quad 0 \leq p_i \leq p_i^{max} \tag{5.85}$$

$$\forall i, \quad 0 \leq u_i \leq u_i^{max} \quad \text{or} \quad w_i^{min} \leq u_i \leq w_i^{max} \tag{5.86}$$

The following remarks are made:

- p_m is the IL reserve offered at the minimum time in order to satisfy the reserve requirement, Eq. 5.81. This value cannot be greater than the amount dispatched at the minimum time, Eq. 5.84, where the set Q_E are all the IL offers where $t_{i,p} = t_{min}$.
- Q_B, the set of IL offers initiated before the minimum time, $t_{i,p} < t_{min}$.
- W_B, a set of SR offers that finishes ramping before the minimum time, $t_{i,u} < t_{min}$ and $t_{i,u} + u_i/g_i \leq t_{min}$. The latter of these conditions is expressed in Eq. 5.86 as $u_i \leq w_i^{max} = g_i(t_{min} - t_{i,u})$.

- W_T, a set of SR offers that continue ramping past the minimum time, $t_{i,u} < t_{min}$ and $t_{min} \leq t_{i,u} + u_i/g_i$. In Eq. 5.86, $w_i^{min} = g_i(t_{min} - t_{i,u}) \leq u_i$. The total argument in the summation over $i \in W_T$ in both Eq. 5.81 and 5.82 is a constant.

- In the minimum frequency constraint, Eq. 5.82, the frequency limit is found, $f_{lim} = f_{lim}(t_{min})$. The frequency level, $f_k = f_{lim}$, if $k > 1$ then the frequency limit constraints for $j < k$ in Eq. 5.83 are not required. Because it is known that these constraints will be satisfied regardless.

- $Q_{B,j}$, the set of IL offers initiated before the step change in frequency limit, $t_{i,p} < t_{j,l}$.

- $W_{B,j}$, a set of SR offers that stop ramping before the step change, $t_{i,u} < t_{j,l}$ and $t_{i,u} + u_i/g_i \leq t_{j,l}$. In Eq. 5.86, $w_i^{max} = g_i(t_{j,l} - t_{i,u})$.

- $W_{T,j}$ a set of SR offers that continue ramping past the step change, $t_{i,u} < t_{j,l}$ and $t_{j,l} \leq t_{i,u} + u_i/g_i$. In Eq. 5.86, $w_i^{min} = g_i(t_{j,l} - t_{i,u})$.

- The sets Q_B and $Q_{B,j}$ share a hierarchy, $Q_{B,j} \subseteq Q_{B,j+1}$. For $f_k = f_{lim}$, if $j > k$ then $Q_B \subseteq Q_{B,j}$. However this hierarchy does not form regions, as there is only one possible option for $Q_{B,j}$.

- The sets W_B, W_T, $W_{B,j}$, and $W_{T,j}$ also form a hierarchy. If $i \in W_{T,j}$ and $i \notin W_{T,j+1}$, then $i \in W_{B,j+1}$. As shown already, SR offers divide the feasible space into regions, this is demonstrated by $W_{B,j}$ and $W_{T,j}$ being partitioned different ways from a combined set $W_{B,j} \cup W_{T,j}$.

The optimisation equations for the SR type are:

$$\text{minimise} \quad \sum_i c_{i,p} p_i + \sum_i c_{i,u} u_i \tag{5.87}$$

$$\text{subject to} \quad \sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i(t_{min} - t_{i,u}) = R \tag{5.88}$$

$$\sum_{i \in Q_B} p_i t_{i,p} + \sum_{i \in W_B} (u_i t_{i,u} + \frac{u_i^2}{2g_i}) + \sum_{i \in W_T} \frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) \leq -2Hf_{lim} \tag{5.89}$$

$$\forall j, \quad -Rt_{j,l} + \sum_{i \in Q_{B,j}} p_i(t_{j,l} - t_{i,p}) + \sum_{i \in W_{B,j}} \left(u_i(t_{j,l} - t_{i,u}) - \frac{u_i^2}{2g_i}\right) + \sum_{i \in W_{T,j}} \frac{g_i}{2}(t_{j,l} - t_{i,u})^2 \geq 2Hf_j \tag{5.90}$$

$$t_{min}^{min} \leq t_{min} \leq t_{min}^{max} \tag{5.91}$$

$$\forall i, \quad 0 \leq p_i \leq p_i^{max} \tag{5.92}$$

$$\forall i, \quad 0 \leq u_i \leq u_i^{max} \quad \text{or} \quad w_i^{min} \leq u_i \leq w_i^{max} \tag{5.93}$$

The main difference between IL type and SR type is that p_m is added as a variable for the first, and t_{min} is added as a variable for the second. There is also one SR region that does not have to be solved at all, this is when W_T is empty, because the entire space of this region can be found in the other regions. More precisely in the situation where W_T is empty, in this region, there has to exist at least one $i \in W_B$ where $u_i = g_i(t_{min} - t_{i,u})$, based on the definition of minimum time in Eq. 5.9 and 5.10. Since there exists, i, then the new region by $W_B \leftarrow W_B \setminus \{i\}$ and $W_T \leftarrow \{i\}$ holds a part of the original region because $u_i = g_i(t_{min} - t_{i,u})$ is an element of the new. If the critical SR offer changes in the original region, i.e. i changes, then another new region can be found on the boundary. Eventually the whole region is partitioned to the others.

A formula is desired to show for the number of IL offers, number of SR offers, and the number of frequency limit step changes, the maximum number of possible regions. The worst possible configuration of critical times, in avoiding a large number of regions, is shown in Figure 5.15. The most regions are gained by having all the frequency limit step changes within the period of SR ramping. Next all the IL offers are placed after the SR offers start ramping and before the frequency limits. Also assume the first feasible minimum time is within $t_{1,u}$ and $t_{2,u}$, and the last feasible minimum time is between $t_{Nu-1,e}^{max}$ and $t_{Nu,e}^{max}$.

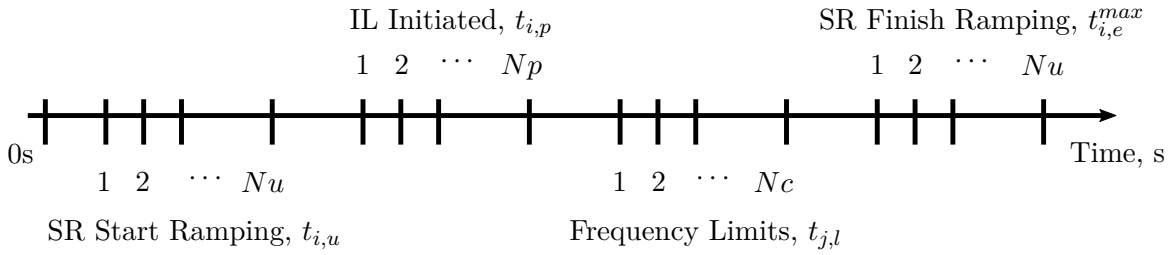


Figure 5.15 The worst case configuration of critical times to give the most regions.

The total number of regions by the four time periods are:

1. While the SR offers start ramping.

$$= \sum_{i=1}^{Nu} (Nc+1)^{Nu-i} \Big((Nc+2)^i - 1 \Big)$$

$$= (Nc+2)\Big((Nc+2)^{Nu} - (Nc+1)^{Nu} \Big) - \frac{1}{Nc}\Big((Nc+1)^{Nu} - 1 \Big)$$

$$= (Nc+2)^{Nu+1} - \frac{1}{Nc}\Big((Nc+1)^{Nu+2} - 1\Big) \tag{5.94}$$

2. While the IL offers are initiated.

$$= Np(Nc+2)^{Nu} + Np\Big((Nc+2)^{Nu} - 1\Big) \tag{5.95}$$

3. While passing the step changes in frequency limit.

$$= \sum_{i=1}^{Nc} \Big((Nc+2-i)^{Nu} - 1\Big) \tag{5.96}$$

4. While the SR offers stop ramping.

$$= \sum_{i=1}^{Nu-1} (2^{Nu-i} - 1) = 2^{Nu} - Nu - 1 \tag{5.97}$$

Approximating the total from the first two periods, i.e. the positive terms of Eq. 5.94 and 5.95:

$$Nr \approx (2Np + Nc + 2)(Nc+2)^{Nu} \tag{5.98}$$

If $Nc = 3$, $Nu = 10$, and $Np = 5$, the approximation (Eq. 5.98) gives a total of 146,484,375 regions, the actual number is 142,001,624 based on the number from each Eq. 5.94-5.97. Clearly it is impractical to solve for every region. If each solution required 1 second to solve, it would take roughly four and a half years to find the optimal solution.

5.3.2 Region Convexity

For each region to be solved by a QCP solver, it is necessary for it to be convex. This is not strictly implied from the previous section on proving the convexity of the whole space, as the size of the space has changed for both the IL and SR type regions. The space for the IL type is defined by the vector $\mathbf{v} = [H, R, p_m, p_1, \cdots, p_{Np}, u_1, \cdots, u_{Nu}]$, and for the SR type by $\mathbf{v} = [H, R, t_{min}, p_1, \cdots, p_{Np}, u_1, \cdots, u_{Nu}]$. Since p_m and t_{min} are not part of the general frequency limit equations, the convexity of these equations is not repeated, as they have been proved from Section 5.2.2. It is the reserve requirement and minimum frequency constraints that need to be checked.

Starting with the IL type problem, for Eq. 5.81, 5.84, 5.85, and 5.86, they are linear and convexity can be inferred directly. For Eq. 5.82, to prove convexity is to show for all $\lambda \in [0, 1]$ that $\mathbf{v} = (1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2$ also satisfies the equation, where both $\mathbf{v}_1$ and $\mathbf{v}_2$ are solutions. Create the function:

$$F_m(\mathbf{v}) = -2Hf_{lim} - p_m t_{min} - \sum_{i \in Q_B} p_i t_{i,p} - \sum_{i \in W_B} (u_i t_{i,u} + \frac{u_i^2}{2g_i}) - \sum_{i \in W_T} \frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) \quad (5.99)$$

It needs to be shown that $F_m((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2) \geq 0$, given $F_m(\mathbf{v}_1) \geq 0$ and $F_m(\mathbf{v}_2) \geq 0$. Like in Section 5.2.2, it is easier to show $F_m((1-\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2) - (1-\lambda)F_m(\mathbf{v}_1) - \lambda F_m(\mathbf{v}_2) \geq 0$, as each term can be analysed individually. The inertial constant, H, IL dispatch, p_i for $i \in Q_B$, and including p_m, these terms are all equal to zero as reiterated from Eq. 5.28 to 5.30. Moving to the SR offers, for $i \in W_B$:

$$\begin{aligned} &-\Big((1-\lambda)u_{i,1} + \lambda u_{i,2}\Big)t_{i,u} - \frac{\Big((1-\lambda)u_{i,1} + \lambda u_{i,2}\Big)^2}{2g_i} + (1-\lambda)\Big(u_{i,1}t_{i,u} + \frac{u_{i,1}^2}{2g_i}\Big) + \lambda\Big(u_{i,2}t_{i,u} + \frac{u_{i,2}^2}{2g_i}\Big) \\ &= \frac{1}{2g_i}\Big(-(1-\lambda)^2 u_{i,1}^2 - 2\lambda(1-\lambda)u_{i,1}u_{i,2} - \lambda^2 u_{i,2}^2 + (1-\lambda)u_{i,1}^2 + \lambda u_{i,2}^2 \Big) \\ &= \lambda(1-\lambda)\frac{(u_{i,1} - u_{i,2})^2}{2g_i} \geq 0 \end{aligned} \quad (5.100)$$

For $i \in W_T$, the summation of like terms is zero:

$$-\frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) + (1-\lambda)\frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) + \lambda\frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) = 0 \quad (5.101)$$

Now for the SR type problem, create the equivalent function for the minimum

frequency constraint:

$$F_m(\mathbf{v}) = -2Hf_{lim} - \sum_{i \in Q_B} p_i t_{i,p} - \sum_{i \in W_B} (u_i t_{i,u} + \frac{u_i^2}{2g_i}) - \sum_{i \in W_T} \frac{g_i}{2}(t_{min}^2 - t_{i,u}^2) \quad (5.102)$$

The important difference between Eq. 5.102 and 5.99 is that t_{min} is now a variable. For $i \in W_T$:

$$-\frac{g_i}{2}\left(\Big((1-\lambda)t_{min,1} + \lambda t_{min,2}\Big)^2 - t_{i,u}^2\right) + (1-\lambda)\frac{g_i}{2}(t_{min,1}^2 - t_{i,u}^2) + \lambda\frac{g_i}{2}(t_{min,2}^2 - t_{i,u}^2)$$

$$= \lambda(1-\lambda)\frac{g_i(t_{min,1} - t_{min,2})^2}{2} \geq 0 \quad (5.103)$$

Hence both the IL and SR type regions are convex.

5.3.3 Region Tracking

This section presents an algorithm to move from region to region to find the global minimum. The algorithm is based on the principle that if the local minimum falls on a boundary, then the local minimum of the other regions on the same boundary will be less than or equal to the first. The rules of the algorithm are first presented in an example before generalising them.

Consider a three dimensional problem with two SR offers, and one IL offer. The main parameters of the offers are shown in Table 5.1. The two SR offers start ramping first, and then IL offer reacts later. The risk is 400 MW, and there is frequency limit step change at 4.5 seconds. The level of the frequency limits themselves are not important for defining each region, and since it is very difficult to draw them in a three dimensional figure, they are omitted for now. In Figure 5.16 the structure of all the regions are seen, in Figure 5.17 and 5.18 they are individually separated to better view and isolate them.

Table 5.1 Parameters for two SR offers and one IL offer example problem.

Offer	$t_{i,u}/t_{i,p}$ s	g_i MW s^{-1}	u_i^{max}/p_i^{max} MW
SR 1	0.5	40	250
SR 2	4/3	60	250
IL 1	3	-	250

All feasible regions are above the plane defined by the reserve requirement constraint: reserve $\geq$ risk, Eq. 5.7. The vertices of the reserve requirement plane at the boundaries of the reserve limits are marked by the numbers (1) to (6) in Figure 5.16. Building the structure of the regions starts by finding the first time the minimum frequency can occur, this is at 3 seconds, where the IL offer can satisfy the reserve requirement with the earlier contributions from the two SR offers, therefore the first set of regions are of the IL type. The boundary to these regions is the plane $p_1 = 200$, the IL type regions on the side $p_1 \geq 200$ form the first set of regions with the earliest minimum frequency. At 3 seconds, SR 1 and SR 2 can both provide 100 MW, therefore 200 MW from IL 1 is needed to cover the risk of 400 MW. There are nine IL type regions in the space $p_1 \geq 200$, these are labelled from (a) to (i) as shown in Figure 5.17. The division amongst these regions is determined by when the two SR offers can stop ramping.

If the amount of IL 1 dispatched is less than 200 MW, then the minimum frequency delays so that one of the SR offers is critical in satisfying the reserve requirement, and the next set of regions are of the SR type. The rest of the feasible space is divided into three main sections: SR 1 is determining the minimum time, SR 2 is setting the minimum time, and both SR 1 and 2 are together. The boundary between these three

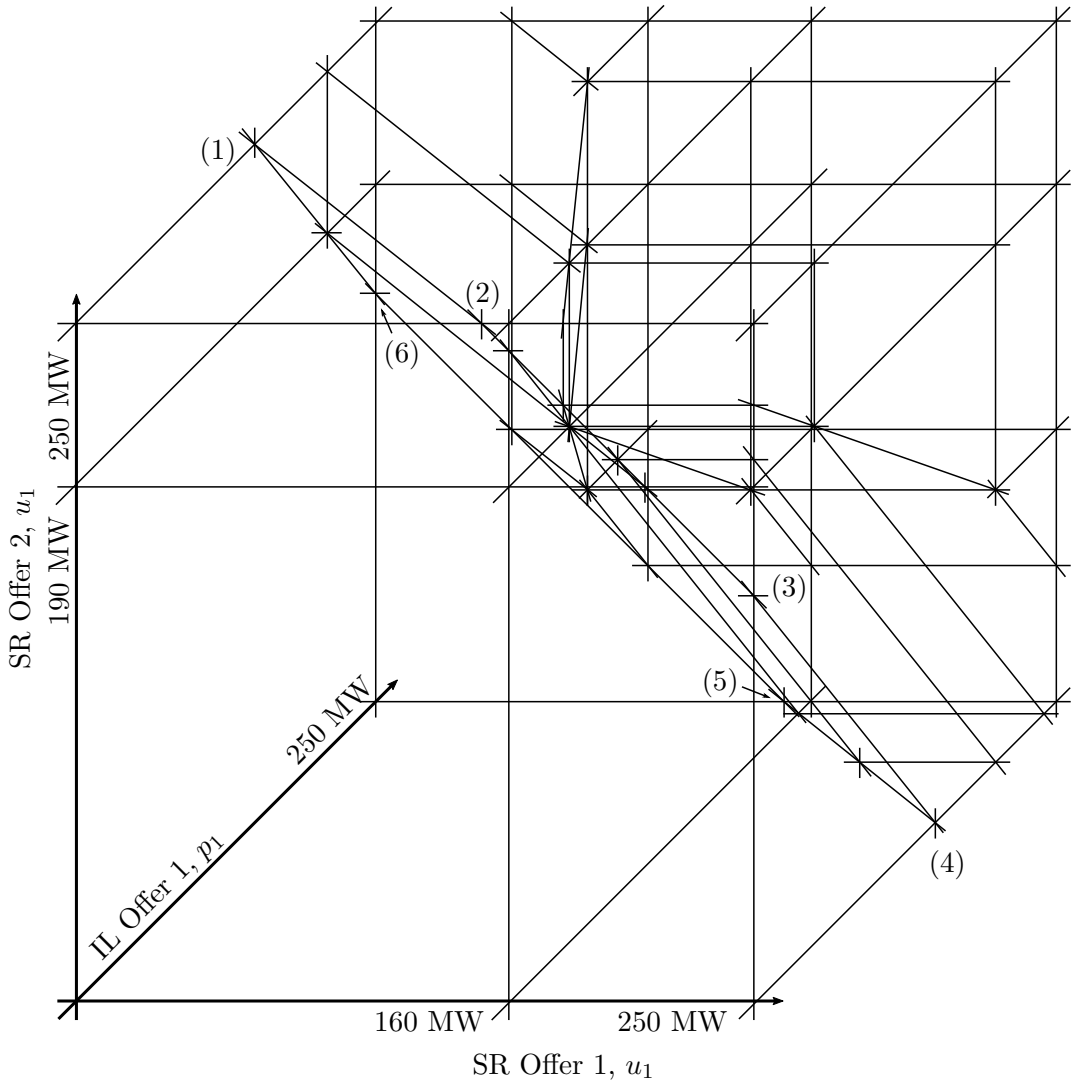


Figure 5.16 Division of the reserve offer space into different regions, for the problem of two SR offers and one IL offer.

sections is the line:

$$\mathbf{v}(t) = \begin{pmatrix} u_1 \\ u_2 \\ p_1 \end{pmatrix} = \begin{pmatrix} 40(t-0.5) \\ 60(t-1.33) \\ 500-100t \end{pmatrix} \tag{5.104}$$

This line is along the reserve requirement plane, which is more closely shown in Figure 5.19, where the points (1) to (6) of Figure 5.19 match those of Figure 5.16. The line connects the point (100,100,200) to (180,220,0) as shown in Figure 5.19, passing through (160,190,50) where both SR offers continue ramping past the step change in frequency limit at 4.5 seconds. The three sections are defined as:

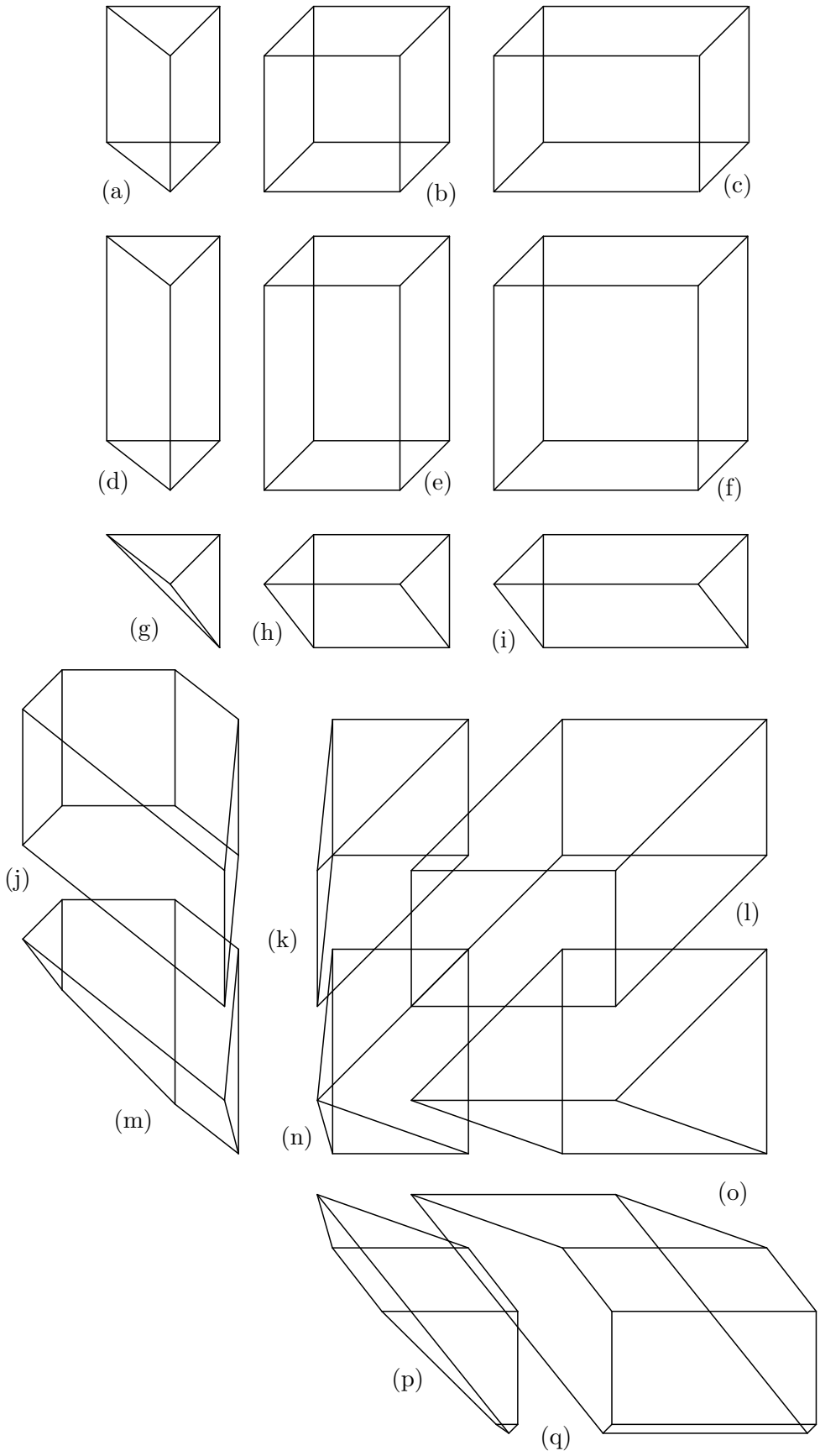


Figure 5.17 Expanded view of each region, for the example of two SR offers and one IL offer.

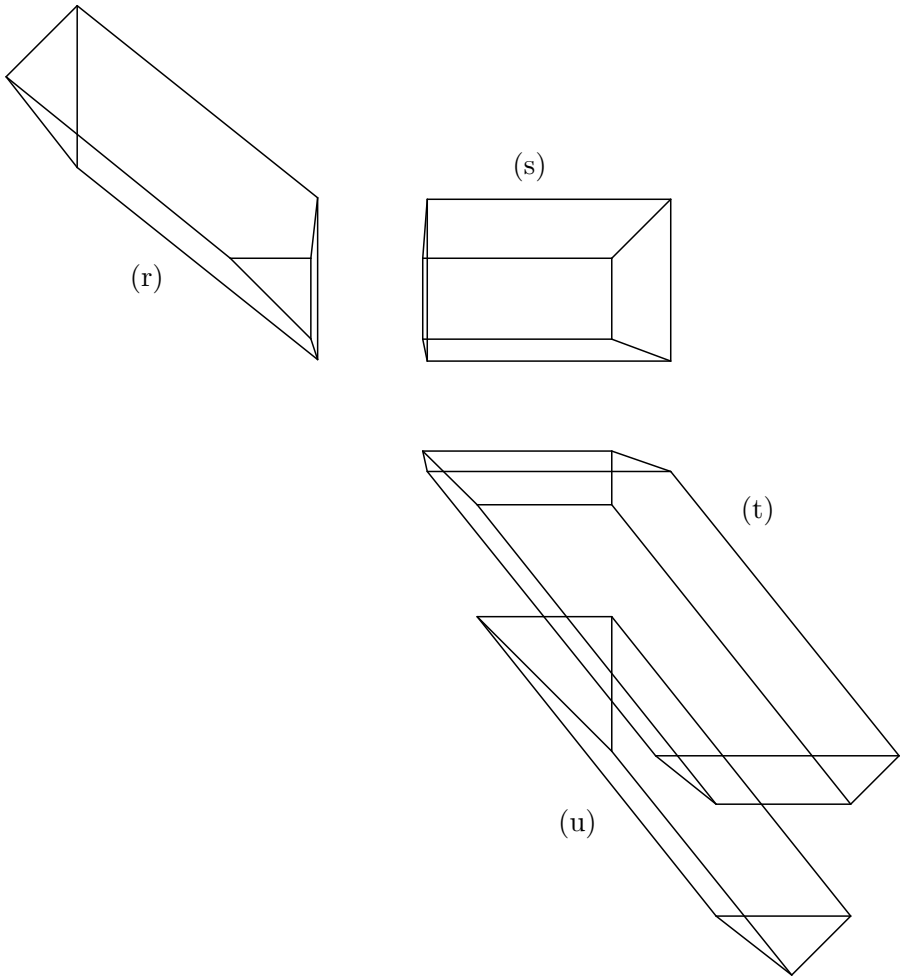


Figure 5.18 Second set of expanded regions.

1. SR 1 sets minimum time. $u_1 \geq 40(t_{min} - 0.5)$, $u_2 \leq 60(t_{min} - 1.33)$, $t_{min} \geq 3$. The time of the minimum frequency, $t_{min} = 0.5 + (400 - u_2 - p_1)/40$. Includes regions (p), (q), (t), and (u) as shown in Figures 5.17 and 5.18.

2. SR 2 sets minimum time. $u_1 \leq 40(t_{min} - 0.5)$, $u_2 \geq 60(t_{min} - 1.33)$, $t_{min} \geq 3$. $t_{min} = 1.33 + (400 - u_1 - p_1)/60$. Includes regions (j), (m), and (r) as shown in Figures 5.17 and 5.18.

3. Both SR 1 and 2. $u_1 \geq 40(t_{min} - 0.5)$, $u_2 \geq 60(t_{min} - 1.33)$. $t_{min} = (500 - p_1)/100 \geq 3$. Includes regions (k), (l), (n), (o), and (s) as shown in Figures 5.17 and 5.18.

The main sections are divided again depending on whether the SR offers continue ramping past the step change in frequency limit or not at $t_{1,l} = 4.5$ seconds. For SR 1 offer this occurs when $u_1 = 160$, and for SR 2 when $u_2 = 190$. Also, since SR 1 finishes ramping after SR 2 finishes ramping, there is an extra region (u) when the minimum time is after $t_{2,e}^{max} = 5.5$ seconds. The details of all the regions are listed in Table 5.2, so

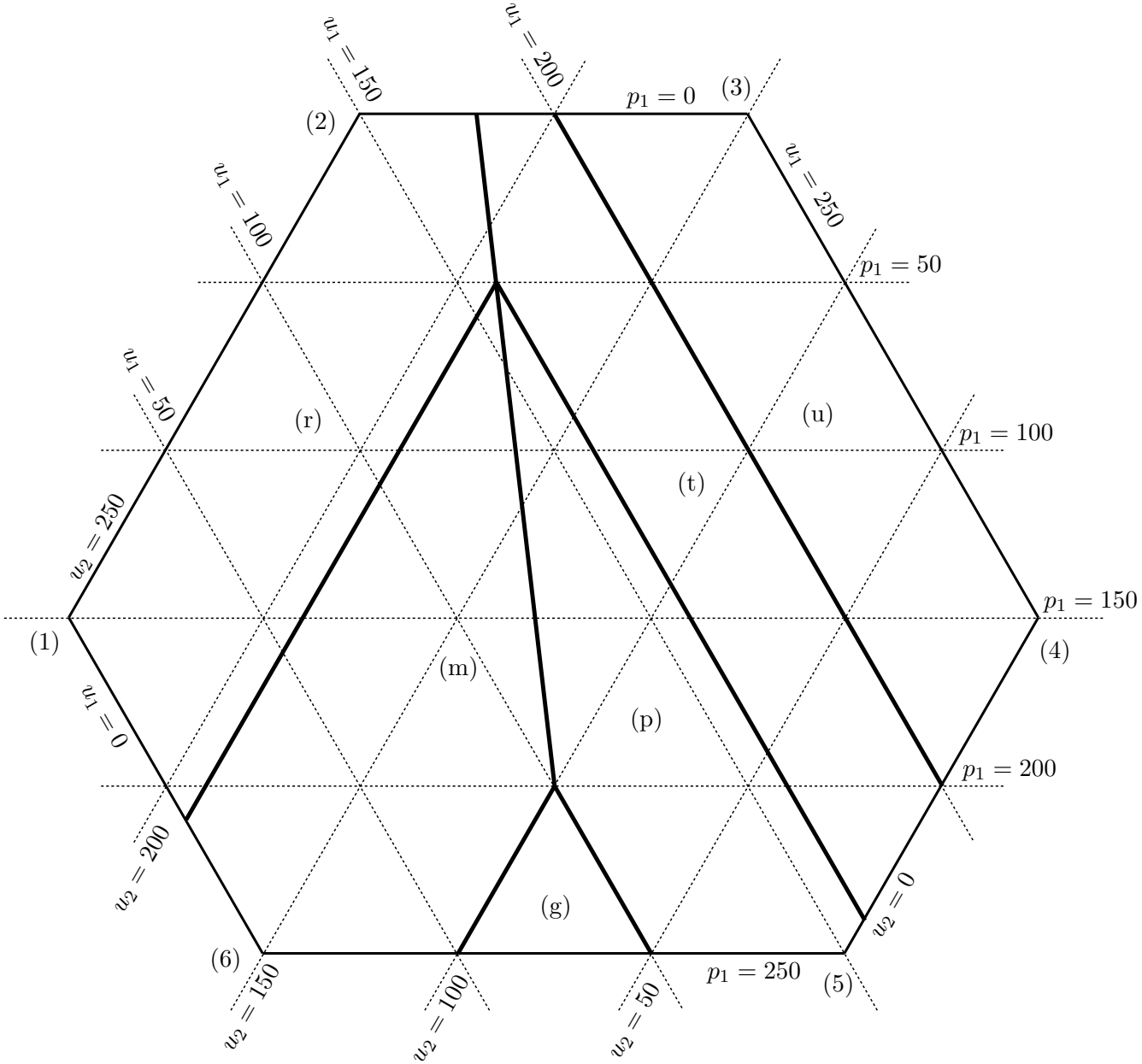


Figure 5.19 A slice of the feasible solution space where the total reserve is equal to the risk, the slice is cut from Figure 5.16

that the formulations of Eqs. 5.80-5.86 and Eqs. 5.87-5.93 for IL and SR type regions respectively can be made.

Table 5.2 Details of each region, for the Q and W sets the number refers to which IL and SR offers respectively.

	Type	t_{min}	Q_B	Q_E	W_B	W_T	$W_{B,1}$	$W_{T,1}$	Label
1	IL	3.0	-	1	1,2	-	1,2	-	(g)
2	IL	3.0	-	1	1	2	1,2	-	(d)
3	IL	3.0	-	1	1	2	1	2	(a)
4	IL	3.0	-	1	2	1	1,2	-	(h)
5	IL	3.0	-	1	2	1	2	1	(i)
6	IL	3.0	-	1	-	1,2	1,2	-	(e)
7	IL	3.0	-	1	-	1,2	1	2	(b)
8	IL	3.0	-	1	-	1,2	2	1	(f)
9	IL	3.0	-	1	-	1,2	-	1,2	(c)
10	SR	3.0 - 4.5	1	-	2	1	1,2	-	(p)
11	SR	3.0 - 4.5	1	-	2	1	2	1	(q)
12	SR	4.5 - 5.5	1	-	2	1	-	-	(t)
13	SR	5.5 - 6.75	1	-	2	1	-	-	(u)
14	SR	3.0 - 4.5	1	-	1	2	1,2	-	(m)
15	SR	3.0 - 4.5	1	-	1	2	1	2	(j)
16	SR	4.5 - 5.5	1	-	1	2	-	-	(r)
17	SR	3.0 - 4.5	1	-	-	1,2	1,2	-	(n)
18	SR	3.0 - 4.5	1	-	-	1,2	1	2	(k)
19	SR	3.0 - 4.5	1	-	-	1,2	2	1	(o)
20	SR	3.0 - 4.5	1	-	-	1,2	-	1,2	(l)
21	SR	4.5 - 5.5	1	-	-	1,2	-	-	(s)

There are a total of 21 possible regions. The approximation of Eq. 5.98 would suggest there is a possibility of 45, and following the more accurate Eq. 5.94 to 5.97 would suggest there is a possibility of 33. The discrepancy between 21 and 33 is because there are no feasible SR type regions for t_{min} before 3 seconds, and so total from Eq. 5.94 should not be counted, even though the configuration of critical times follows the worst case scenario. Therefore the total possible number of regions from the three equations, Eq. 5.95, 5.96, and 5.97 is 21 in accordance with the number obtained.

To become more familiar with the different region types the optimisation equations are shown for two regions. For the IL type, consider region (a); the corresponding objective function from Eq. 5.80 is:

$$\text{minimise} \quad c_{1,p}p_1 + c_{1,u}u_1 + c_{2,u}u_2 \tag{5.105}$$

subject to the following constraints, the first of which defines p_m from Eq. 5.81:

$$p_m + u_1 + 60(3 - 1.33) = 400$$

$$\Rightarrow p_m + u_1 = 300 \tag{5.106}$$

The minimum frequency constraint from Eq. 5.82 is:

$$3p_m + 0.5u_1 + \frac{u_1^2}{80} + \frac{60}{2}(3^2 - 1.33^2) \leq 2Hf_0$$

$$\Rightarrow 3p_m + 0.5u_1 \leq 2Hf_0 - 216.67 \tag{5.107}$$

The frequency limit from Eq. 5.83 is:

$$-400 \times 4.5 + p_1(4.5 - 3) + u_1(4.5 - 0.5) - \frac{u_1^2}{80} + \frac{60}{2}(4.5 - 1.33)^2 \geq 2Hf_1$$

$$\Rightarrow 1.5p_1 + 4u_1 - \frac{u_1^2}{80} \geq 2Hf_1 + 1499.17 \tag{5.108}$$

The limits on p_m, Eq. 5.84, and p_1, Eq. 5.85, are:

$$0 \leq p_m \leq p_1 \quad \text{and} \quad 0 \leq p_1 \leq 250 \tag{5.109}$$

Lastly, the limits on the two SR offers come in the form $0 \leq u_1 \leq g_1(t_{min} - t_{1,u})$ and $g_2(t_{1,l} - t_{2,u}) \leq u_2 \leq u_2^{max}$ from Eq. 5.86, and are simplified to:

$$0 \leq u_1 \leq 100 \quad \text{and} \quad 190 \leq u_2 \leq 250 \tag{5.110}$$

The top and bottom of the triangular prism of region (a) in Figure 5.17 is created by $190 \leq u_2 \leq 250$. The right side plane is $u_1 \leq 100$, the back side plane is $p_1 \leq 250$, and the plane from the left at an angle is from the substitution of Eq. 5.106 into $p_m \leq p_1$ from Eq. 5.109. Since for this region the dispatch of u_2 is restricted to stop ramping after the time of the minimum frequency, 3 seconds, and the time of step change, 4.5 seconds, it is seen that the variable u_2 has no influence on the minimum frequency constraint, Eq. 5.107, and the frequency limit, Eq. 5.108. This is because the contribution of u_2 to the constraints remains unchanged in the range, $190 \leq u_2 \leq 250$. Therefore if the price of SR offer 2, $c_{2,u}$, is positive the optimal solution will have $u_2 = 190$, and if negative $u_2 = 250$, and lastly if zero then there is not a unique solution, and any u_2 in the range would be valid.

The second region to consider is region (t) of SR type. The objective from Eq. 5.87:

$$\text{minimise} \quad c_{1,p}p_1 + c_{1,u}u_1 + c_{2,u}u_2 \tag{5.111}$$

subject to the following constraints, the definition of t_{min} in Eq. 5.88 is:

$$p_1 + 40(t_{min} - 0.5) + u_2 = 400$$

$$\Rightarrow p_1 + 40t_{min} + u_2 = 420 \tag{5.112}$$

The minimum frequency constraint from Eq. 5.89 is:

$$3p_1 + \frac{40}{2}(t_{min}^2 - 0.5^2) + 1.33u_2 + \frac{u_2^2}{120} \geq 2Hf_1$$

$$\Rightarrow 3p_1 + 20t_{min}^2 + 1.33u_2 + \frac{u_2^2}{120} \leq -2Hf_1 + 5 \tag{5.113}$$

The range of possible times for t_{min} comes from Eq. 5.91, and the limit on p_1 from Eq. 5.92, as shown in following equations:

$$4.5 \leq t_{min} \leq 5.5 \quad \text{and} \quad 0 \leq p_1 \leq 250 \tag{5.114}$$

Finally, the limits on the two SR offers are $g_1(t_{min} - t_{1,u}) \leq u_1 \leq u_1^{max}$ and $0 \leq u_2 \leq g_2(t_{min} - t_{2,u})$ and are expressed with numeric values in the following equations:

$$40(t_{min} - 0.5) \leq u_1 \leq 250 \quad \text{and} \quad 0 \leq u_2 \leq 60(t_{min} - 1.33) \tag{5.115}$$

The influence of each equation on the boundary of the feasible space is shown in Figure 5.20 with the isolated view of region (t). The frequency limit, Eq. 5.90, at 4.5 seconds is not included among the equations because the minimum frequency constraint, Eq. 5.113, implies that the frequency limit is always satisfied as $t_{1,l} \leq t_{min}^{min}$. For $t \leq t_{min}$ it is known that $f(t) \geq f_1$, therefore the frequency at 4.5 seconds is always greater than or equal to f_1.

An example has been described, now to develop an algorithm to solve for the global optimum. Assume for now that a QCP solver exists to find the local minimum in each region and can determine if it is on a boundary or not. This solver is discussed in the next section. Also let the inertia, H, and risk, R, be constant for the rest of the chapter.

The algorithm starts by finding a region which has a feasible space, but at the moment it is not known whether any feasible points exist. Therefore the first step is to check if the point $(u_1^{max}, u_2^{max}, p_1^{max})$ satisfies the reserve requirement, the minimum frequency constraint, and the frequency limit. If it does then region (c) has feasible points, if not then the problem has no solution and cannot continue further.

The second step is to find the local minimum of region (c); mark the local minimum in the variables by the asterisk $(*)$. The local minimum can be found in several different possible places as shown in Figure 5.21 depending on the choice of H, f_0, f_1, and

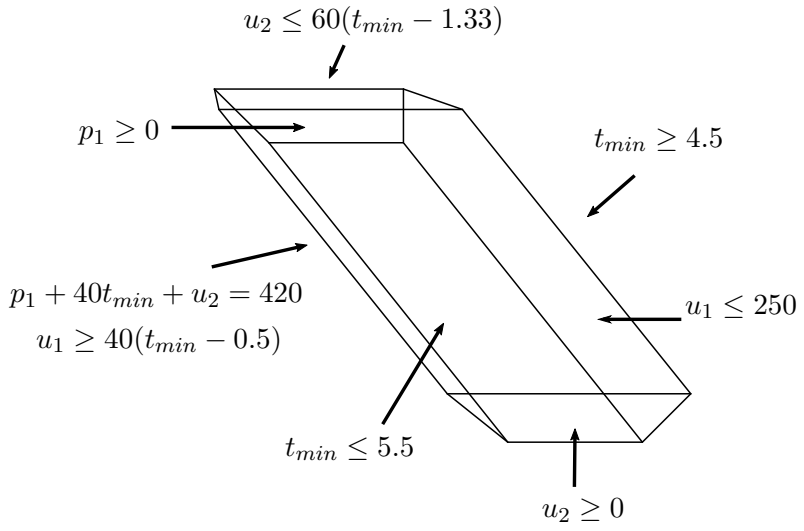


Figure 5.20 The boundaries of region (t) and the associated equations.

prices. The other parameters have already been specified. The next step depends on the position of the local minimum, there are three possibilities:

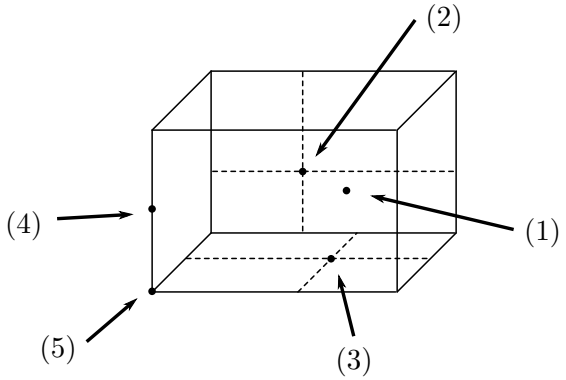


Figure 5.21 Possible positions of local minimums in region (c).

1. The local minimum does not fall on any boundaries, (1), or it falls on a boundary that does not connect with any other region, (2), as seen in Figure 5.21. Then the global minimum has been found and is the local minimum. To prove this, it is shown by contradiction: assume there is a point $\mathbf{v}_2$ not in region (c) and is the global minimum. Let $\mathbf{v}_1$ be the local minimum in region (c). Since the entire feasible space is convex, a straight line can be drawn between the two points and still be inside the solution space: for all $\lambda \in [0, 1]$ every point $\mathbf{v} = (1 - \lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2$ is feasible. The price vector is $\mathbf{c}$ and the total cost is $\mathbf{c}^T\mathbf{v}$. Since $\mathbf{v}_2$ is the global minimum $\mathbf{c}^T\mathbf{v}_2 < \mathbf{c}^T\mathbf{v}_1$, there is a point on the boundary of region (c) at $\lambda = \lambda_b$, where $\mathbf{v}_b = (1 - \lambda_b)\mathbf{v}_1 + \lambda_b\mathbf{v}_2$, the cost of which is:

$$\mathbf{c}^T\Big((1-\lambda_b)\mathbf{v}_1+\lambda_b\mathbf{v}_2\Big)=\mathbf{c}^T\mathbf{v}_1+\lambda_b(\mathbf{c}^T\mathbf{v}_2-\mathbf{c}^T\mathbf{v}_1)<\mathbf{c}^T\mathbf{v}_1$$

This is a contradiction because a point on the boundary, $\mathbf{v}_b$, has a cost less than the local minimum, $\mathbf{v}_1$, a contradiction. Therefore the local minimum, $\mathbf{v}_1$, has to be the global minimum, and the process is finished.

2. The local minimum shares a boundary with one other region, point (3) in Figure 5.21. The boundary is shared with region (f), the next step is to find the local minimum of (f), because the local minimum of (f) less than or equal in cost to that of (c). The location of the local minimum in (f) will determine the next step.

   If the local minimum of (f) happens to be the same as point (3) from region (c), then point (3) is the global minimum. This is because if it is not the global minimum, then a contradiction can be found where there is either a point in (c) or (f) with a lower cost than the shared local minimum. That is a repetition of the proof from the previous option with some modifications.

3. The local minimum is on multiple boundaries which it shares with multiple regions. For example, point (4) borders with (b), (k), and (l), whereas point (5) includes those of (4) and (e), (f), (n), and (o). The difficulty is which region should be checked next? There is not necessarily a correct answer to this question, as the global minimum can be found more quickly picking different regions under the right conditions. However a good policy, if all the prices are positive, is to pick the one directly opposite of (c) through the point of the local minimum. In this case region (n) if the local minimum is found on point (5), or (k) if the local minimum is like point (4).

The algorithm transitions from region to region until it has found the global minimum. However it might find a position where a conditional decision needs to be made. For example, if the local minimum of region (n) is found on the tip of its spike, i.e. the (160, 190, 50) coordinate. This point borders with (k), (l), (o), (j), (m), (p), (q), (r), (s), and (t). There is not an obvious region to choose next, as the desired choice would be to transition across the reserve requirement boundary, but this is not possible as this is outside the feasible space. The next best option is to choose between (r), (m), (p), and (t), the options that have a main border with the reserve requirement constraint, as seen in Figure 5.19. To determine which one to check next, project the negative price vector, $-\mathbf{c}$, onto the reserve requirement plane, the direction it points is the next option. If it points along a boundary, e.g. the line $u_2 = 200$ between (r) and (m), then randomly pick one. If the projection does not point in any direction, then the point (160, 190, 50) is one global minimum, there may be more depending on whether other points on the reserve requirement plane are feasible.

5.3.4 The Region Tracking Algorithm

So far an example is shown of tracking from one region to another. To generalise the process, both IL and SR type regions are reformulated to remove p_m and t_{min} respectively. This is so that both types can be placed in the same space, i.e. with the components $\mathbf{v} = [p_1, \cdots, p_{Np}, u_1, \cdots, u_{Nu}]^T$ assuming inertia and risk are constant. Once this is done, an algorithm is presented to transition from one region to another. It is possible to develop a solving methodology with inertia and risk as variables, but to simplify the algorithm it is assumed they are constant.

Starting with IL type regions, the equality constraint, Eq. 5.81, is used to derive an equation for p_m:

$$p_m = R - \sum_{i \in Q_B} p_i - \sum_{i \in W_B} u_i - \sum_{i \in W_T} g_i(t_{min} - t_{i,u}) \tag{5.116}$$

since $t_{min} = t_{i,p}$ from those $i \in Q_E$, it is not necessary to remove t_{min} as it will be for SR type regions. The constraints on p_m from Eq. 5.84 then become:

$$R \geq \sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i(t_{min} - t_{i,u}) \tag{5.117}$$

$$R \leq \sum_{i \in Q_E} p_i + \sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i(t_{min} - t_{i,u}) \tag{5.118}$$

The IL and SR offer limits, Eq. 5.85 and 5.86, do not change as they are not dependent on p_m. The frequency limit and minimum frequency constraint equations are not shown because they do not form the boundary between regions, rather they form the outer boundary. For the SR type problem, t_{min} is derived from Eq. 5.88 and is equal to:

$$t_{min} = \frac{R - \sum_{i \in Q_B} p_i - \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i t_{i,u}}{\sum_{i \in W_T} g_i} \tag{5.119}$$

The limits on t_{min}, Eq. 5.91, are rearranged as:

$$\sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i(t_{min}^{min} - t_{i,u}) \leq R \tag{5.120}$$

$$\sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i + \sum_{i \in W_T} g_i(t_{min}^{max} - t_{i,u}) \geq R \tag{5.121}$$

Also some of the SR offers may be limited by t_{min}, e.g. $u_i \leq g_i(t_{min} - t_{i,u})$ or

$u_i \geq g_i(t_{min} - t_{i,u})$, i.e. Eq. 5.93. The limit for $u_k \leq g_k(t_{min} - t_{k,u})$ where $k \in W_B$:

$$\sum_{i \in Q_B} p_i + \sum_{i \in W_B \setminus \{k\}} u_i \quad + \Big(1 + \frac{\sum_{i \in W_T} g_i}{g_k}\Big) u_k \leq R + \sum_{i \in W_T} g_i t_{i,u} - t_{k,u} \Big(\sum_{i \in W_T} g_i \Big) \quad (5.122)$$

and for $u_k \geq g_k(t_{min} - t_{k,u})$ where $k \in W_T$:

$$\sum_{i \in Q_B} p_i + \sum_{i \in W_B} u_i \quad + \Big(1 + \frac{\sum_{i \in W_T \setminus \{k\}} g_i}{g_k}\Big) u_k \geq R + \sum_{i \in W_T \setminus \{k\}} g_i t_{i,u} - t_{k,u} \Big(\sum_{i \in W_T \setminus \{k\}} g_i \Big) \quad (5.123)$$

Now that the boundaries have been expressed in the desired space, a distinction is made between boundaries that form the outside of the entire feasible space, and the boundaries that form the regions inside it. The distinction is made by specifying the outside boundaries and the rest are then inside boundaries, the outside boundaries are:

1. The minimum and maximum reserve limits: $0 \leq p_i \leq p_i^{max}$ and $0 \leq u_i \leq u_i^{max}$.

2. The reserve requirement constraint, Eq. 5.7:

$$\sum_{i \in Q} p_i + \sum_{i \in W} u_i \geq R \quad (5.124)$$

   where $Q = \{1, 2, \cdots, Np\}$ and $W = \{1, 2, \cdots, Nu\}$, i.e. all the offers. These sets are related to the others by $Q_B \cup Q_E \subseteq Q$ and $W_B \cup W_T \subseteq W$.

3. The frequency limit and minimum frequency constraints.

The inside boundaries are those of Eq. 5.117, 5.118, 5.120, 5.121, 5.122, and 5.123. However there are special situations when Eq. 5.118 and 5.123 are the same as the reserve requirement constraint, Eq. 5.124, and is counted as an outside boundary in these situations. The conditions for this to occur are:

1. For the IL type region Eq. 5.118, W_T has to be empty and $W_B = W$. For example, region (g) has a boundary like this, as seen in Figure 5.19 by having a face on the reserve requirement plane.

2. For the SR type Eq. 5.123, W_T has to have a single element k, $W_B = W \setminus \{k\}$, and $Q_B = Q$. Regions (r), (m), (p), (t), and (u) lie on the reserve requirement boundary, which is also seen in Figure 5.19.

The final step before describing the algorithm is to simplify the inside boundary constraints. The inside boundaries are linear inequalities, therefore they are expressed in vector notation:

$$\mathbf{a}_i \bullet \mathbf{v} \leq b_i \tag{5.125}$$

where i is a general index and does not refer to reserve offers, $\mathbf{a}_i$ are the coefficients and b_i is a constant. For example consider the boundary $t_{min} \leq 5.5$ for region (t), this corresponds to Eq. 5.121, $p_1 + u_2 + 40(5.5 - 0.5) \geq 400$, and in the vector notation:

$$\begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} \bullet \begin{pmatrix} u_1 \\ u_2 \\ p_1 \end{pmatrix} \leq -200 \tag{5.126}$$

It is convenient to define the region as an intersection of two spaces: the entire feasible space, S_f, and the space defined by the inside boundary constraints, Z_r, where r is an index for each region. So that feasible space for a region is $\Pi_r = S_f \cap Z_r$. The space defined by the inside boundaries is:

$$Z_r = \{\mathbf{v} \in \mathbb{R}^N \quad \text{s.t.} \quad \forall i \leq Nb(r), \quad \mathbf{a}_{i,r} \bullet \mathbf{v} \leq b_{i,r}\} \tag{5.127}$$

where $Nb(r)$ is the number of inside boundary constraints for region r, and $N = Np + Nu$. The algorithm to solve the optimisation is shown in flow chart of Figure 5.22 and is annotated in the following steps:

Algorithm 5.1

1. Find the region where $(p_1^{max}, \cdots, p_{Np}^{max}, u_1^{max}, \cdots, u_{Nu}^{max})^T$ is on the corner.
2. Check if this point is feasible in the region. If it is not feasible, then there is no solution to the whole optimisation problem, and stop here.
3. Find the local minimum for the region. The next step depends on the position of the local minimum.
   (a) The local minimum, $\mathbf{v}^*$, has no internal boundaries where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$, then stop, the local minimum for this region is the global minimum for the whole space.
   (b) If this is not the first time step 3 has been executed, if the current local minimum is the same as the previous local minimum, check if the current local minimum is the global minimum by considering which outside boundaries it is on. Otherwise continue to Step 4.

(c) Else continue.

4. Find the next region from the last local minimum. From the internal boundaries the local minimum is on, i.e. i where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$, determine the next region, where these boundaries in the new region are $-\mathbf{a}_{i,r} \bullet \mathbf{v} \leq -b_{i,r}$ for the same i the local minimum is on. This is what happens for options (3), (4), and (5) in Figure 5.21 in choosing the next region. Now go back to step 3 and repeat.

This is the basic form of the algorithm, there are other complexities not expressed here, but in Appendix B the detailed algorithm is presented, and proved that for a feasible problem it will terminate, so that a solution will always be found.

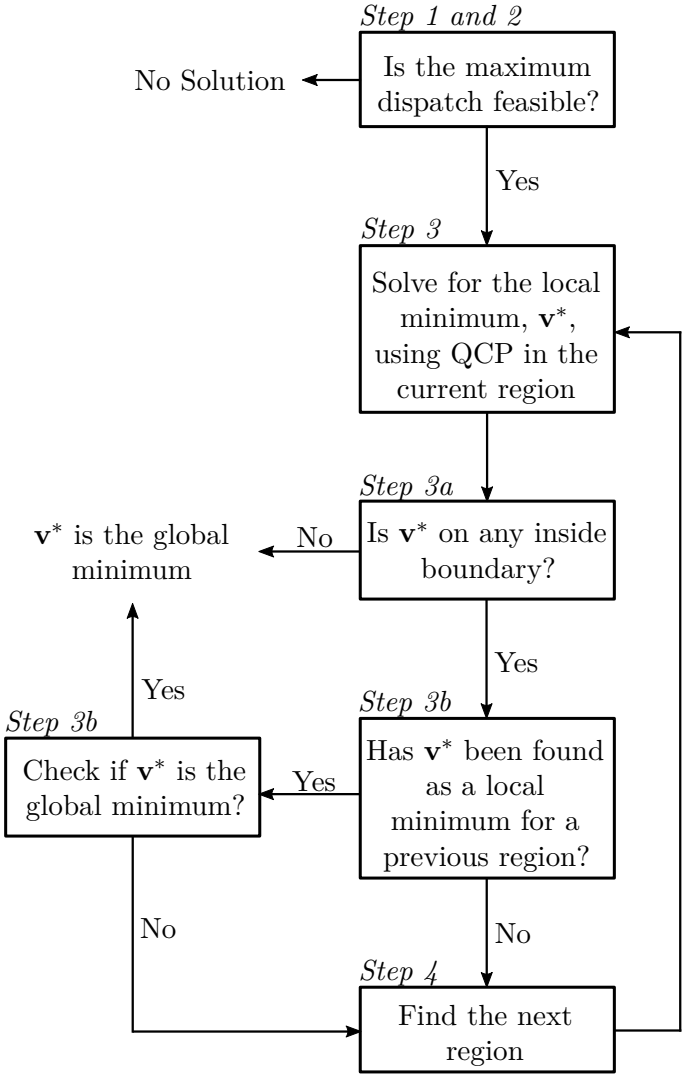


Figure 5.22 Flow chart of the solving methodology to find the global minimum of the optimization problem

The Number of Region Optimisations

The amount of time required for this algorithm to execute is dependent on the number of regions optimised. When co-optimising with the energy market, one optimisation of a region is like optimising once the entire electricity market in New Zealand's SPD formulation. It has already been shown that it is impractical to solve every region, now the algorithm provides a much more practical number.

It is not easy to provide an exact number of the regions that have to be optimised, but rather an estimate of the maximum required is given. It is left to Section 5.4 by way of examples to recommend a likely number, and therefore the more important value.

The algorithm starts in one corner, $\mathbf{v} = (p_1^{max}, \cdots, p_{Np}^{max}, u_1^{max}, \cdots, u_{Nu}^{max})^T$, of the entire feasible space. It transitions in the direction of the minimum cost, and finds the global minimum on the edge of the feasible space. Therefore the maximum number of iterations is the distance between the starting region and furthest away one. Consider again Figure 5.15, assuming it is only possible to transition from one time range to another at each step, then the region furthest away is at the end of the time line, as the first region is always at the start of the time line. The maximum number of iterations required is then estimated to be:

$$Nr \approx 2(Np + Nu) + Nc - 1 \tag{5.128}$$

This is significantly more practical as it is linearly dependent upon the size of the problem, as compared to Eq. 5.98 with the exponential term $(Nc+2)^{Nu}$. Following the example of that section, where $Nc = 3$, $Nu = 10$, and $Np = 5$, the maximum number of regions required to be solved is 32, i.e. 32 seconds compared to four and a half years. The actual number of regions, depending on the problem, is going to be considerably less than 32.

5.3.5 Quadratically Constrained Programming

This section describes the optimisation process: transforming Eq. 5.80 to 5.93 into a standard form, necessary conditions for local minima, and determining which boundaries the minimum is on. Details of the solver itself are not provided as convex QCP is a developed field of research. A simplified solver is built in MATLAB. It is based on solving the necessary conditions for optimality via a form of Newton's Method, which provides greater accuracy in the solution after each iteration. For the process of jumping from one feasible region to another to work, a method of determining which boundary the local minimum is on is required. This section focuses on this methodology.

Transforming to a Standard Form

The goal is to transform Eqs. 5.80 to 5.86 and 5.87 to 5.93 into a form that is consistent for both IL and SR type problems. This form is:

$$\text{minimise} \quad \mathbf{c}^T\mathbf{y} \tag{5.129}$$

$$\text{subject to} \quad G\mathbf{y} + H\mathbf{y}^2 + \mathbf{m} = 0 \tag{5.130}$$

$$\text{and} \quad \mathbf{y} \geq 0 \tag{5.131}$$

$\mathbf{y}$ are the variables, $\mathbf{c}$ is the price vector, and G, H, and $\mathbf{m}$ are the coefficients of the quadratic constraints, the first two are matrices. The sense of $\mathbf{y}^2$ is the element-wise squaring so that $\mathbf{y}^2$ is also a vector of the same size as $\mathbf{y}$. This form is not the most general formulation of quadratically constrained problems, but since no variables are cross multiplied, e.g. there are no $t_{min}p_i$ terms or any other combinations, it is easier to use this form.

The first step in the transformation process is converting to a vector form. The second step is to bring all terms to the left hand side. For an IL type problem, the result of transforming Eq. 5.80-5.86 is:

$$\text{minimise} \quad \mathbf{c}_p^T\mathbf{p} + \mathbf{c}_u^T\mathbf{u} \tag{5.132}$$

$$\text{subject to} \quad p_m + \mathbf{a}^T\mathbf{p} + \mathbf{b}^T\mathbf{u} + m_1 = 0 \tag{5.133}$$

$$p_m\mathbf{t}_m + A\mathbf{p} + B\mathbf{u} + D\mathbf{u}^2 + \mathbf{m}_2 \leq 0 \tag{5.134}$$

$$p_m - \mathbf{d}^T\mathbf{p} \leq 0 \tag{5.135}$$

$$\mathbf{p} - \mathbf{p}^{max} \leq 0 \tag{5.136}$$

$$\mathbf{w}^{min} - \mathbf{u} \leq 0 \quad \text{and} \quad \mathbf{u} - \mathbf{w}^{max} \leq 0 \tag{5.137}$$

$$\text{plus} \quad p_m \geq 0, \quad \mathbf{p} \geq 0, \quad \text{and} \quad \mathbf{u} \geq 0 \tag{5.138}$$

where p_m, $\mathbf{p}$, and $\mathbf{u}$ are the variables. $\mathbf{c}_p$ and $\mathbf{c}_u$ are the price vectors for IL and SR offers respectively. $\mathbf{a}$, $\mathbf{b}$, and m_1 are

$$a_i = \begin{cases} 1 & i \in Q_B \\ 0 & i \notin Q_B \end{cases} \quad \text{and} \quad b_i = \begin{cases} 1 & i \in W_B \\ 0 & i \notin W_B \end{cases} \tag{5.139}$$

$$m_1 = -R + \sum_{i \in W_T} g_i(t_{min} - t_{i,u}) \tag{5.140}$$

The coefficients of the frequency constraints are ordered with the first row being the minimum frequency constraint, Eq. 5.82, and the subsequent rows are all the other frequency limits, Eq. 5.83. Therefore the first element of $\mathbf{t}_m$ is t_{min} and all the subsequent elements are zero. The other coefficients are:

$$A_{i,j} = \begin{cases} t_{j,p} & i = 1, \quad j \in Q_B \\ -(t_{i-1,l} - t_{j,p}) & i > 1, \quad j \in Q_{B,i-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.141}$$

$$B_{i,j} = \begin{cases} t_{j,u} & i = 1, \quad j \in W_B \\ -(t_{i-1,l} - t_{j,u}) & i > 1, \quad j \in W_{B,i-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.142}$$

$$D_{i,j} = \begin{cases} 1/2g_j & i = 1, \quad j \in W_B \\ 1/2g_j & i > 1, \quad j \in W_{B,i-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.143}$$

$$m_{2,i} = \begin{cases} 2Hf_{lim} + \sum_{j \in W_T} \frac{g_j}{2}(t_{min}^2 - t_{j,u}^2) & i = 1 \\ 2Hf_{i-1} + Rt_{i-1,l} - \sum_{j \in W_{T,i-1}} \frac{g_j}{2}(t_{i-1,l} - t_{j,u})^2 & i > 1 \end{cases} \tag{5.144}$$

The indices, i and j, are the rows and columns respectively. Their precise meaning is not entirely consistent with previous sections, where i referred to offers and j to frequency limits. In the Eqs. 5.141-5.144, i refers to frequency constraint, and j to an offer, this is an almost reversal of meaning. For the upper limit on p_m, Eq. 5.135, the

coefficient $\mathbf{d}$ is:

$$d_i = \begin{cases} 1 & i \in Q_E \\ 0 & i \notin Q_E \end{cases} \tag{5.145}$$

The rest of the coefficients should be self explanatory. The third step is to make Eqs. 5.134 to 5.137 equality constraints by the introduction of slack variables. The last step is then to combine all the equations to form G, H, and $\mathbf{m}$. Therefore the vector of variables $\mathbf{y}$ and the price vector $\mathbf{c}$ are:

$$\mathbf{y} = \begin{bmatrix} p_m \\ \mathbf{p} \\ \mathbf{u} \\ \mathbf{s}_f \\ s_m \\ \mathbf{s}_p \\ \mathbf{s}_u^{min} \\ \mathbf{s}_u^{max} \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 0 \\ \mathbf{c}_p \\ \mathbf{c}_u \\ \mathbf{0}_{Ns\times 1} \end{bmatrix} \tag{5.146}$$

where $Ns = Nc + Np + 2Nu + 2$ is the number of slack variables. The slack variables, $\mathbf{s}_f$, is for Eq. 5.134; the other slack variables are for the other equations, Eqs. 5.135 to 5.137. The constraint matrices are:

$$G = \begin{bmatrix} 1 & \mathbf{a}^T & \mathbf{b}^T & \mathbf{0}_{1\times Nc} & 0 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nu} \\ \mathbf{t}_m & A & B & I_{Nc} & \mathbf{0}_{Nc\times 1} & \mathbf{0}_{Nc\times Np} & \mathbf{0}_{Nc\times Nu} & \mathbf{0}_{Nc\times Nu} \\ 1 & -\mathbf{d}^T & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nc} & 1 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nu} \\ \mathbf{0}_{Np\times 1} & I_{Np} & \mathbf{0}_{Np\times Nu} & \mathbf{0}_{Np\times Nc} & \mathbf{0}_{Np\times 1} & I_{Np} & \mathbf{0}_{Np\times Nu} & \mathbf{0}_{Np\times Nu} \\ \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & -I_{Nu} & \mathbf{0}_{Nu\times Nc} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & I_{Nu} & \mathbf{0}_{Nu\times Nu} \\ \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & I_{Nu} & \mathbf{0}_{Nu\times Nc} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & \mathbf{0}_{Nu\times Nu} & I_{Nu} \end{bmatrix} \tag{5.147}$$

$$H = \begin{bmatrix} \mathbf{0}_{1\times(Np+1)} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Ns} \\ \mathbf{0}_{Nc\times(Np+1)} & D & \mathbf{0}_{Nc\times Ns} \\ \mathbf{0}_{(1+Np+2Nu)\times(Np+1)} & \mathbf{0}_{(1+Np+2Nu)\times Nu} & \mathbf{0}_{(1+Np+2Nu)\times Ns} \end{bmatrix} \tag{5.148}$$

$$\mathbf{m} = \begin{bmatrix} m_1 \\ \mathbf{m}_2 \\ 0 \\ -\mathbf{p}^{max} \\ \mathbf{w}^{min} \\ -\mathbf{w}^{max} \end{bmatrix} \tag{5.149}$$

where I_{Nc} is a square identity matrix with Nc rows and columns, I_{Np} and I_{Nu} are identity matrices as well.

More quickly, the standardisation of the SR type problem is now repeated similarly to the IL type problem. The constraints in the vector form are:

$$\text{minimise} \quad \mathbf{c}_p^T \mathbf{p} + \mathbf{c}_u^T \mathbf{u} \tag{5.150}$$

$$\text{subject to} \quad g_m t_{min} + \mathbf{a}^T \mathbf{p} + \mathbf{b}^T \mathbf{u} + m_1 = 0 \tag{5.151}$$

$$t_{min}^2 \mathbf{g}_m + A\mathbf{p} + B\mathbf{u} + D\mathbf{u}^2 + \mathbf{m}_2 \leq 0 \tag{5.152}$$

$$t_{min}^{min} - t_{min} \leq 0 \quad \text{and} \quad t_{min} - t_{min}^{max} \leq 0 \tag{5.153}$$

$$\mathbf{p} - \mathbf{p}^{max} \leq 0 \tag{5.154}$$

$$t_{min}\mathbf{w}^{low} + \mathbf{w}^{min} - \mathbf{u} \leq 0 \quad \text{and} \quad \mathbf{u} - t_{min}\mathbf{w}^{high} - \mathbf{w}^{max} \leq 0 \tag{5.155}$$

$$\text{plus} \quad t_{min} \geq 0, \quad \mathbf{p} \geq 0, \quad \text{and} \quad \mathbf{u} \geq 0 \tag{5.156}$$

where

$$g_m = \sum_{i \in W_T} g_i \quad \text{and} \quad m_1 = -R - \sum_{i \in W_T} g_i t_{i,u}$$

The first element of $\mathbf{g}_m$ is $g_m/2$, and all the other elements are zero as t_{min} does not feature in the other frequency limits. The matrices A, B, and D remain the same, and $\mathbf{m}_2$ remains mostly the same except for the first element, where:

$$m_{2,1} = 2Hf_{lim} - \frac{1}{2} \sum_{i \in W_T} g_i t_{i,u}^2$$

The coefficients of the SR offer limits are:

$$\mathbf{w}^{low} = \begin{cases} g_i & \text{Boundary of the type} \quad g_i(t_{min} - t_{i,u}) \leq u_i \\ 0 & \text{otherwise} \end{cases} \tag{5.157}$$

$$\mathbf{w}^{min} = \begin{cases} -g_i t_{i,u} & \text{Boundary of the type} \quad g_i(t_{min} - t_{i,u}) \leq u_i \\ w_i^{min} & \text{otherwise} \end{cases} \tag{5.158}$$

$$\mathbf{w}^{high} = \begin{cases} g_i & \text{Boundary of the type} \quad u_i \leq g_i(t_{min} - t_{i,u}) \\ 0 & \text{otherwise} \end{cases} \tag{5.159}$$

$$\mathbf{w}^{max} = \begin{cases} -g_i t_{i,u} & \text{Boundary of the type} \quad u_i \leq g_i(t_{min} - t_{i,u}) \\ w_i^{max} & \text{otherwise} \end{cases} \tag{5.160}$$

The parameters of the standard form are:

$$\mathbf{y} = \begin{bmatrix} t_{min} \\ \mathbf{p} \\ \mathbf{u} \\ \mathbf{s}_f \\ \mathbf{s}_t \\ \mathbf{s}_p \\ \mathbf{s}_u^{min} \\ \mathbf{s}_u^{max} \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 0 \\ \mathbf{c}_p \\ \mathbf{c}_u \\ \mathbf{0}_{Ns\times 1} \end{bmatrix} \tag{5.161}$$

$$G = \begin{bmatrix} g_m & \mathbf{a}^T & \mathbf{b}^T & \mathbf{0}_{1\times Nc} & 0 & 0 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nu} \\ \mathbf{0}_{Nc\times 1} & A & B & I_{Nc} & \mathbf{0}_{Nc\times 1} & \mathbf{0}_{Nc\times 1} & \mathbf{0}_{Nc\times Np} & \mathbf{0}_{Nc\times Nu} & \mathbf{0}_{Nc\times Nu} \\ -1 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nc} & 1 & 0 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nu} \\ 1 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nc} & 0 & 1 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Nu} \\ \mathbf{0}_{Np\times 1} & I_{Np} & \mathbf{0}_{Np\times Nu} & \mathbf{0}_{Np\times Nc} & \mathbf{0}_{Np\times 1} & \mathbf{0}_{Np\times 1} & I_{Np} & \mathbf{0}_{Np\times Nu} & \mathbf{0}_{Np\times Nu} \\ \mathbf{w}^{low} & \mathbf{0}_{Nu\times Np} & -I_{Nu} & \mathbf{0}_{Nu\times Nc} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & I_{Nu} & \mathbf{0}_{Nu\times Nu} \\ -\mathbf{w}^{high} & \mathbf{0}_{Nu\times Np} & I_{Nu} & \mathbf{0}_{Nu\times Nc} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times 1} & \mathbf{0}_{Nu\times Np} & \mathbf{0}_{Nu\times Nu} & I_{Nu} \end{bmatrix} \tag{5.162}$$

$$H = \begin{bmatrix} 0 & \mathbf{0}_{1\times Np} & \mathbf{0}_{1\times Nu} & \mathbf{0}_{1\times Ns} \\ \mathbf{g}_m & \mathbf{0}_{Nc\times Np} & D & \mathbf{0}_{Nc\times Ns} \\ \mathbf{0}_{(2+Np+2Nu)\times 1} & \mathbf{0}_{(2+Np+2Nu)\times Np} & \mathbf{0}_{(2+Np+2Nu)\times Nu} & \mathbf{0}_{(2+Np+2Nu)\times Ns} \end{bmatrix} \tag{5.163}$$

$$\mathbf{m} = \begin{bmatrix} m_1 \\ \mathbf{m}_2 \\ t_{min}^{min} \\ -t_{min}^{max} \\ -\mathbf{p}^{max} \\ \mathbf{w}^{min} \\ -\mathbf{w}^{max} \end{bmatrix} \tag{5.164}$$

where $Ns = Nc + Np + 2Nu + 3$.

Necessary Conditions for Optimum

The local minimum is found by transforming the optimisation problem into a set of equations. These are called the Karush-Kuhn-Tucker conditions. For the standard form of Eqs. 5.129 to 5.131 these are:

$$\mathbf{c} + \left(G^T + 2H^T \otimes [\mathbf{y}]_{Ne}\right)\boldsymbol{\lambda} - \boldsymbol{\mu} = 0 \tag{5.165}$$

$$\text{subject to} \quad G\mathbf{y} + H\mathbf{y}^2 + \mathbf{m} = 0 \tag{5.166}$$

$$\forall i \quad y_i \mu_i = 0 \tag{5.167}$$

$$\mathbf{y} \geq 0 \quad \text{and} \quad \boldsymbol{\mu} \geq 0 \tag{5.168}$$

where the expression $[\mathbf{y}]_{Ne}$ is a matrix by repeating $\mathbf{y}$ as columns, i.e. $[\mathbf{y}]_{Ne} = [\mathbf{y}, \mathbf{y}, \cdots, \mathbf{y}]$, the number of repetitions is determined by Ne, the number of rows in G or the number of equality constraints in the problem. The operator, $\otimes$, is the element-wise multiplication of two matrices, as opposed to the normal matrix multiplication. The vector $\boldsymbol{\lambda}$ is the KKT multipliers for the equality constraints of Eq. 5.130, and $\boldsymbol{\mu}$ is the KKT multipliers for the inequality constraints of Eq. 5.131.

The method of solving these equations can be done by an interior point like method. An initial guess is made of the solution for $\mathbf{y}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ in the quadrant determined by Eq. 5.168; successive steps are taken via Newtons Method to solve Eqs. 5.165 to 5.167 more accurately upon each iteration. If the Newton step would shift $\mathbf{y}$ and $\boldsymbol{\mu}$ outside the feasible quadrant then the step is shortened. The method stops when the size of the step is below an error tolerance. There are other considerations to be made but these have been dealt with in the literature.

Determining when on a Boundary

For the region tracking algorithm to work, it has to be known when the local minimum is on a boundary. A naive approach to doing this would be to say for example, if for a SR offer $u_i = 0$ then it is on a boundary or for a more complicated example, Eq. 5.135, $p_m = \mathbf{d}^T\mathbf{p}$, which is more easily determined when the slack variable $s_m = 0$, then it is also on a boundary. Therefore a boundary is determined whenever $y_i = 0$. However, this approach is not adequate because through using an interior point like method there is always a slight error in the solution for $\mathbf{y}$, and never quite equals zero. A less naive approach is developed to determine when the local minimum is on a boundary.

The goal is to determine when $y_i = 0$. The key to making that decision is by

considering Eq. 5.167. The solutions for this equation are when either $y_i = 0$ or $\mu_i = 0$ or both are zero. Therefore the first step in correcting the local minimum is to make a decision for each i on whether $y_i = 0$ or $\mu_i = 0$. The decision is made on whether $y_i \leq \mu_i$ or not, if so then $y_i = 0$ otherwise $\mu_i = 0$. This decision may be incorrect, particularly when both y_i and μ_i are small, therefore an error check is completed to determine if the right decision has been made.

The error check is to solve Eq. 5.165 and 5.166 with a choice of y_i and μ_i equalling zero. Eq. 5.167 does not form a part in this step as this condition has already been satisfied in the decision process. A Newton's method is used to solve these equations, but this time without the boundary requirements of Eq. 5.168. The starting point is obtained from solution already found. The Newton's method stops when a step change is within a tighter error tolerance than the previous solver. If it is found for the solution that $\mathbf{y} \geq 0$ and $\boldsymbol{\mu} \geq 0$ then the correct decision has been made.

It is efficient to simplify Eq. 5.165 and 5.166 in the error checking process. The first step is to rearrange both $\mathbf{y}$ and $\boldsymbol{\mu}$ into zero and non-zero partitions:

$$\left[\frac{\mathbf{y}_v}{\mathbf{0}}\right] = \left[\frac{T_y}{T_\mu}\right]\mathbf{y} \quad \text{and} \quad \left[\frac{\boldsymbol{\mu}_v}{\mathbf{0}}\right] = \left[\frac{T_\mu}{T_y}\right]\boldsymbol{\mu} \tag{5.169}$$

where T_y and T_μ are transformation matrices that reorders the rows of the object it operates on. The transformation of Eq. 5.165 and 5.166 is:

$$\mathbf{c}_\mu + G_\mu^T\boldsymbol{\lambda} - \boldsymbol{\mu}_v = 0 \tag{5.170}$$

$$\mathbf{c}_y + \left(G_y^T + 2H_y^T \otimes [\mathbf{y}_v]_{Ne}\right)\boldsymbol{\lambda} = 0 \tag{5.171}$$

$$G_y\mathbf{y}_v + H_y\mathbf{y}_v^2 + \mathbf{m} = 0 \tag{5.172}$$

where

$$\left[\frac{\mathbf{c}_\mu}{\mathbf{c}_y}\right] = \left[\frac{T_\mu}{T_y}\right]\mathbf{c} \quad \text{and} \quad \left[\begin{array}{c|c} G_y & G_\mu \end{array}\right] = G\left[\begin{array}{c|c} T_y^T & T_\mu^T \end{array}\right]$$
$$\text{and} \quad \left[\begin{array}{c|c} H_y & H_\mu \end{array}\right] = H\left[\begin{array}{c|c} T_y^T & T_\mu^T \end{array}\right] \tag{5.173}$$

From looking at Eqs. 5.170 and 5.171, it is noticed that $\mathbf{y}_v$ and $\boldsymbol{\lambda}$ can be solved firstly through Newton's method and then $\boldsymbol{\mu}_v$ can be solved directly from Eq. 5.170; a computationally faster approach. This simplification should be expected as the decision between whether y_i and μ_i equals zero removes the inequality constraints from the optimisation. Eq. 5.171 and 5.172 are the necessary conditions for optimisation

problems with only equality constraints.

If an incorrect decision is made between the choice of y_i and μ_i being zero, then this will be seen by at least one component of $\mathbf{y}_v$ and $\boldsymbol{\mu}_v$ being less than zero, an infeasible solution. To correct this mistake, at least one of the decisions has to be swapped, maybe more. If $\mathbf{y}$ has N variables then there are $2^N - 1$ possible options. For any reasonably sized problem, $N = 30$, there is roughly one billion options to choose from. It is not practical to try each option systematically until the correct answer is found, therefore a different approach is required.

A good approach is to find the likelihood that a decision was made incorrectly. This is done by creating a variable $z_i = y_i + \mu_i$, where y_i and μ_i are from the first approach of solving Eqs. 5.165 to 5.168. If z_i is small then the likelihood is high, and if large then the likelihood is low. Therefore rank z_i from smallest to largest. The first option is to swap the first pair in the order. If this one fails, then the next one after that, and continue in a binary sequence until a feasible solution is found.

If the initial solution is close then it will not take very many options before the correct solution is found. It is beyond the scope of this research to optimise the solving process, but reducing the number of options to check would be of ideal efficiency.

5.4 SOLVED EXAMPLES

This section demonstrates the working of a solver developed in MATLAB, while highlighting important implications of this formulation. Performance of the solver is analysed by the number of steps required to find the global minimum. The main implication is the general absence of a clear marginal reserve provider, because one provider with faster and more expensive reserve can be dispatched before a cheaper offer. The consequence is a loss of intuitiveness in being able to predict the outcome; the optimisation cannot be approximated by a standard merit-order dispatch based on the offered quantities and prices alone. Therefore reserve providers will find it difficult to adjust their offers in the scheduling process, as adjustments may have undesired effects. However, the positive consequence is that speed of response is valued properly. Although a standard merit-order curve cannot be formulated, a merit-order curve can be produced for when a single frequency limit is the only binding frequency constraint.

Four examples are given to highlight the important implications. Example one has the purpose of going through the solving methodology step by step for the one IL and two SR offer problem of Section 5.3.3. Example two includes more offers, thereby proving the methodology works for larger problems, but also considers the impacts of having and not having additional frequency limits after the initial one. From this example the difficulty of finding the marginal reserve provider is also highlighted. Example three shows a situation where the optimal solution gives the frequency transient at its minimum a discontinuity in its derivative, as a result of an IL offer being critical to satisfying the reserve requirement. Example four analyses what occurs to the total cost of the optimal solution when the inertial constant, H, and the risk, R, which is the largest credible contingency, are varied. The purpose of showing this is to highlight the importance of inertia and risk as variables in the optimisation of reserves, or whether they should be variables at all.

The optimisation form of the examples are:

minimise *reserve availability cost*
subject to *frequency constraints*
and reserve availability limits

Each example cannot be described through one objective function and a single set of simple constraints, as the frequency constraints change form depending on the region in the feasible solution space. Neither is it beneficial to describe the constraints as piecewise functions, as these functions become prohibitively large as the problems get bigger. Neither is it beneficial to describe the problem by listing every region like it is done in Table 5.2, as the number of regions increases quickly with size of the problem, i.e. a combinatorial explosion. The best approach is to simply list the parameters of the

offers and frequency limit function, with the implied understanding that the optimal dispatch is to satisfy keeping the frequency transient above the frequency limit function.

The examples have only considered the smaller problem of optimising reserve against frequency constraints. It has not considered constraints found in common energy optimisations. There are three reasons for this. (1) Focus is primarily placed on understanding the implications of optimising with frequency limits. (2) The solver requires further development to solve problems with mixed integer variables. (3) There are a range of possible energy optimisations, from Unit Commitment type problems to Optimal Power Flow, and to the inclusion of security constraints. Therefore to avoid giving the wrong impression on how far this formulation can be applied, the choice of larger optimisation problem is omitted to retain generality in application.

5.4.1 First Example - Three Offer Problem

The first example is to complete the process of solving the example presented in Section 5.3.3, the problem with one IL offer and two SR offers. Each step in the solving methodology can be analysed and visualised in the feasible solution space. To complete the specification of the problem, the remaining parameters are given appropriate values. Firstly, a price is given to each offer: for SR offer 1 it is \$54/MW, for SR offer 2 it is \$36/MW, and for the IL offer it is \$90/MW. The parameters of each offer are reiterated in Table 5.3.

Table 5.3 Parameters for two SR offers and one IL offer example with prices.

Offer	$t_{i,u}/t_{i,p}$ s	g_i MW s^{-1}	u_i^{max}/p_i^{max} MW	$c_{i,p}/c_{i,u}$ \$/MW
SR 1	0.5	40	250	54
SR 2	4/3	60	250	36
IL 1	3	-	250	90

The initial frequency limit is 48 Hz ($f_0 = -0.04$), and at $t_{1,l} = 4.5$ seconds the frequency limit changes to 48.3 Hz ($f_1 = -0.034$). The inertial constant is set to 15,000 MW s, around the average inertia of the North Island of New Zealand. Lastly, the risk is again 400 MW.

Using the flowchart of Figure 5.22 and Algorithm 5.1, the process of solving this example is reviewed step by step:

1. (Step 1). The maximum dispatch is $(u_1, u_2, p_1) = (250,250,250)$. This point lies on vertex of region (c), as shown in Figure 5.23 by the black dot.

2. (Step 2). The feasibility of (250,250,250) needs to be determined. Starting with the reserve requirement of Eq. 5.7: $u_1 + u_2 + u_3 = 750$ and greater than the risk of 400 MW. Therefore the reserve requirement is satisfied. Determining the feasibility against the minimum frequency constraint, the first reserve to

satisfy the reserve requirement is found: $(w_1, w_2, q_1) = (100,100,200)$. The time of minimum frequency is 3 seconds. Therefore checking against the minimum frequency constraint in Eq. 5.11 via its transformation in Eq. 5.50, which it is reiterated here:

$$\sum_i q_i t_{i,p} + \sum_i (w_i t_{i,u} + \frac{w_i^2}{2g_i}) \leq -2Hf_{lim}(t_{min})$$

$$q_1 t_{1,p} + w_1 t_{1,u} + \frac{w_1^2}{2g_1} + w_2 t_{2,u} + \frac{w_2^2}{2g_2} \leq -2Hf_0$$

$$200 \times 3 + 100 \times 0.5 + \frac{100^2}{2 \times 40} + 100 \times 1.33 + \frac{100^2}{2 \times 60} \leq -2 \times 15,000 \times -0.04$$

the result is the true statement that $991.66 \leq 1200$, with units of MW s, therefore the minimum frequency constraint is satisfied. Since the time of the minimum frequency is before $t_{1,l} = 4.5$ seconds, the only frequency limit of Eq. 5.8 needs to be checked. The total reserve dispatched before 4.5 seconds is $(w_1, w_2, q_1) =$ (160,190,250), checking against the transformed frequency limit of Eq. 5.26 being greater than zero:

$$-2Hf_j - Rt_{j,l} + \sum_i q_i(t_{j,l} - t_{i,p}) + \sum_i (w_i(t_{j,l} - t_{i,u}) - \frac{w_i^2}{2g_i}) \geq 0$$

$$(-2 \times 15,000 \times -0.034) - (400 \times 4.5) + 250 \times (4.5 - 3) + 160 \times (4.5 - 0.5) - \frac{160^2}{2 \times 40}$$

$$+190 \times (4.5 - 1.33) - \frac{190^2}{2 \times 60} \geq 0$$

which gives $215.83 \geq 0$, in units of MW s, therefore the frequency limit is satisfied, and the maximum dispatch is feasible.

3. (Step 3). Optimise region (c), the optimisation problem is:

$$\text{from Eq. 5.80 minimise} \quad 90p_1 + 54u_1 + 36u_2 \tag{5.174}$$

$$\text{subject to the frequency limit of Eq. 5.83} \quad 1.5p_1 - 1179.17 \geq -1020 \tag{5.175}$$

$$\text{and the limits} \quad 200 \leq p_1 \leq 250 \quad 160 \leq u_1 \leq 250 \quad 190 \leq u_2 \leq 250 \tag{5.176}$$

The Eqs. 5.81 and 5.82 are not required to form the problem as they do not contain any variables, as $p_m = 200$ by Eq. 5.81. The frequency limit of Eq. 5.175 can also be omitted, as it is simplified to $p_1 \geq 106.1$, and redundant compared to $p_1 \geq 200$ of Eq. 5.176, which comes from the constraint $p_1 \geq p_m$ in Eq. 5.84. The optimal solution is found by taking the lower limits for each variable in Eq. 5.176: $\mathbf{v}^* = (u_1^*, u_2^*, p_1^*) = (160, 190, 200)$. This local minimum is seen in Figure 5.23 as the first red dot in the sequence.

4. (Step 3a). The local minimum of region (c), $\mathbf{v}^* = (160, 190, 200)$, is on three inside boundaries: $u_1 \geq 160$, $u_2 \geq 190$, and $p_1 \geq 200$. Therefore this point appears not to be the global minimum.

5. (Step 4). The next region to optimise is by determining which region has the following three inside boundaries: $u_1 \leq 160$, $u_2 \leq 190$, and $p_1 \leq 200$. Visually inspecting Figures 5.16 and 5.17, the appropriate region is (n).

6. (Step 3). Optimise region (n), its local minimum is $\mathbf{v}^* = (u_1^*, u_2^*, p_1^*) = (131.6, 151.6, 121.0)$, it is seen by the second red dot in sequence of Figure 5.23.

7. (Step 3a). The local minimum of region (n) is on one inside boundary, $u_1 \geq g_1(t_{min} - t_{1,u})$, and in more detail from Eq. 5.123:

$$p_1 + (1 + g_2/g_1)u_1 \geq R + g_2(t_{2,u} - t_{1,u}) \tag{5.177}$$

Therefore it appears that this point is not the global minimum.

8. (Step 4). The next region to optimise is the one that flips $\geq$ to $\leq$ in Eq. 5.177. From inspecting the position of $(u_1^*, u_2^*, p_1^*) = (131.6, 151.6, 121.0)$ in Figure 5.23, and comparing where that point is next to in other regions in Figures 5.16 and 5.17, the next region to optimise is region (m).

9. (Step 3). Optimise region (m), the local minimum is $\mathbf{v}^* = (u_1^*, u_2^*, p_1^*) = (124, 154, 124.1)$. This seen by the last red dot in the sequence from Figure 5.23.

10. (Step 3a). The local minimum of region (m) is on no inside boundaries, therefore it is the global minimum and the solver now stops. The optimal dispatch in reserve is $p_1 = 124.1$, $u_1 = 124$, and $u_2 = 154$ MW.

For these local minima in regions (c), (n), and (m) the resulting frequency transients can be plotted, Figure 5.24. These frequency transients can be produced from Eq. 5.6. The first line is the frequency transient, labelled by 'Maximum Dispatch' in Figure 5.24, for which all reserve is dispatched, the frequency transient rebounds the quickest. This is not cost efficient, so the solving process relaxes the reserve requirement by finding the

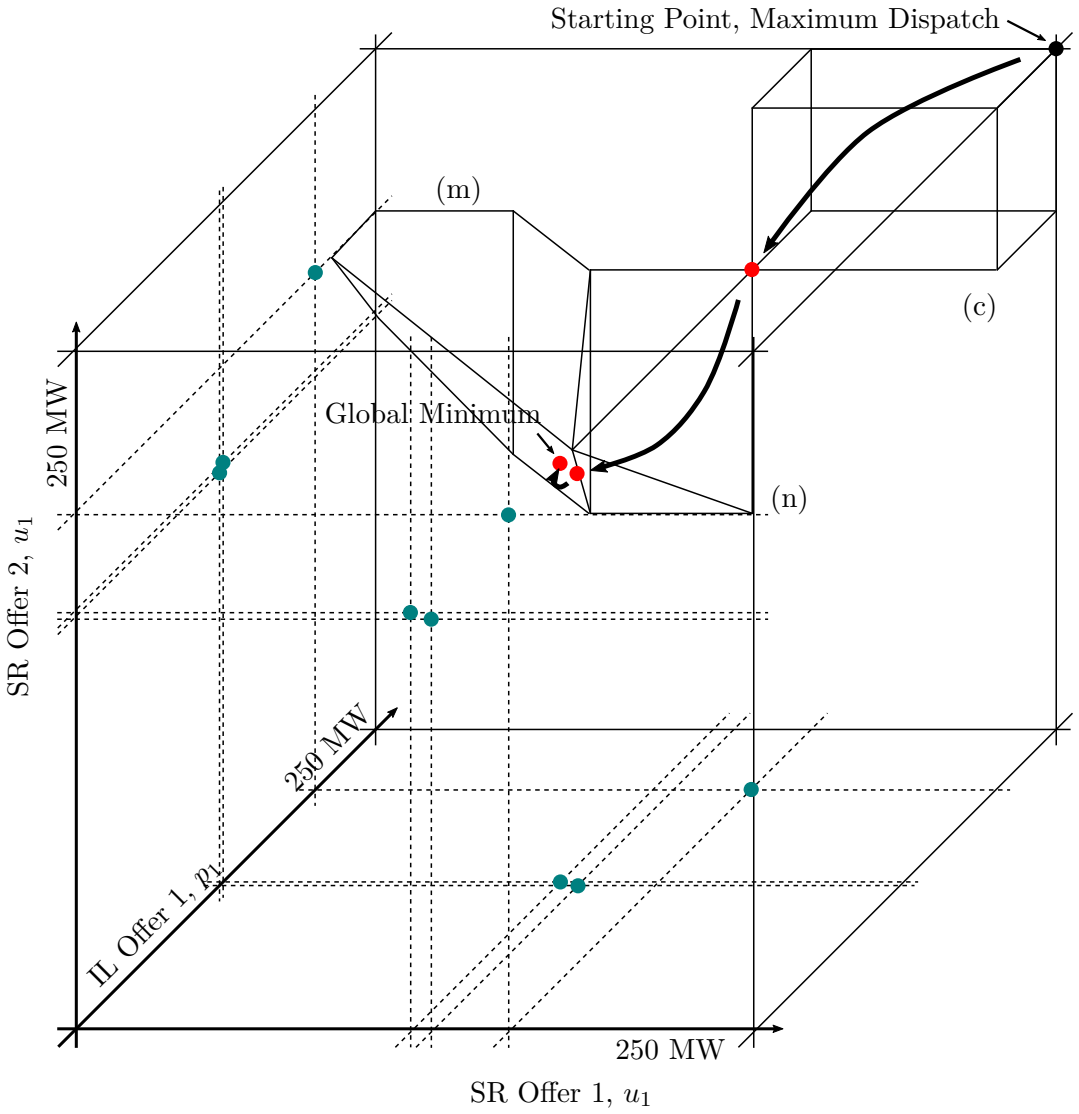


Figure 5.23 Steps through the regions to find the global minimum. The red dots are the local minimums, the teal coloured dots are projections onto the plane.

local optimal of region (c), and continues until it reaches (m). The frequency transient for (m) is restricted by the frequency limit at 4.5 seconds, and it is no longer possible to relax the reserve requirement. Region (n) is also restricted by the same frequency limit, but the difference between the solutions for (m) and (n) is a trade-off for a cheaper reserve dispatch. The difference in what is happening between (c) to (n) and (n) to (m) can be broadly defined into two phases: firstly a reduction phase, and then a trading off phase. This generalisation can be seen in the later examples.

Consider the optimal solution, $(u_1^*, u_2^*, p_1^*) = (124, 154, 124.1)$, it is apparent there is not a single marginal generator, as neither of the three offers are at their minimum or maximum limits. If the optimisation was executed under the usual method without requirements for speed of response and frequency constraints, then SR offer 2 would be dispatched to its maximum, and then SR offer 1 would make the difference, 150 MW, to

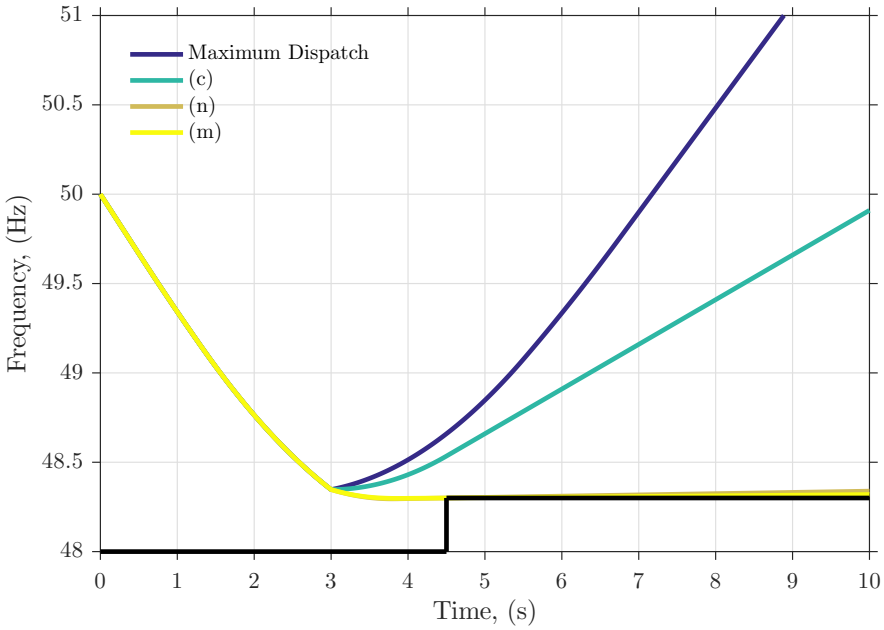


Figure 5.24 Power System frequency transients for each dispatch of reserve in the process of finding the global optimum. First Example.

achieve the optimal solution. The reserve price would then be \$54/MW. However, under the new formulation, in a sense, it is IL offer 1 that is the marginal generator, because for every 1 MW that the risk increases it requires an extra 3 MW of reserve from IL offer 1 to retain the new optimal solution. Therefore following a similar philosophy in reserve pricing, the price is \$270/MW. This is a major difference, the reserve price has increased 5 times. It is not necessarily recommended that this be the pricing methodology, as more thought is required, but is helpful in understanding the trade-offs that are likely to occur in a co-optimised energy and reserves market. That is the decision on how much the risk setter increases its energy dispatch is dependent on whether it displaces generation that has an energy price of \$270/MW or more. The marginal values are derived in Appendix C, as they can not be read directly from the shadow variables of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ in Eqs. 5.165-5.168.

5.4.2 Second Example - Frequency Limits

For this example the performance of the solver is shown for a larger problem, while highlighting the importance of additional frequency limits. It is noticed from the previous example that the optimal solution does not return the frequency to 50 Hz in a reasonable amount of time, as seen by the frequency transient for optimal solution of region (m) in Figure 5.24. This is an unsatisfactory dispatch in order to keep the power system secure. Therefore more frequency limits should be introduced.

It may have been noticed in Figure 5.24 that under this formulation, after the last reserve has finished ramping, the frequency increases linearly indefinitely at the rate determined by the final imbalance between reserve and risk. Therefore the frequency transient is only valid as an accurate model until the last reserve finishes ramping or the time of the last frequency constraint, which ever comes last provided they are both reasonable. Therefore the model only tries to capture the first downwards swing in frequency and rebound. To improve the accuracy of the model, it will be shown that additional frequency limits should be chosen to reflect the likely rebound in frequency from its minimum back to 50 hertz. Therefore the frequency transient after it has returned to 50 hertz in the model is then considered an artefact of the formulation, and will not express any physical implications during an contingency, as subsequent control action will keep the frequency close to 50 hertz.

Table 5.4 List of IL Offers for the Second Example

	Quantity (MW)	Price ($/MW)	$t_{i,p}$ (seconds)
1	10	150	0.9
2	38	126	0.9
3	16	0	1
4	57	65	1
5	42	132	1.1
6	63	118	1.3
7	29	98	1.5
8	50	100	2.5
9	75	50	3
10	18	20	3.3

The example is expressed in the IL and SR offers of Tables 5.4 and 5.5 with a risk of 400 MW and an inertia of 15,000 MW s. The example is firstly solved with only the initial frequency limit of 48 Hz ($f_0 = -0.04$), then the result is compared to the solution with additional frequency limits.

Table 5.5 List of SR Offers for the Second Example

	Quantity (MW)	Price ($/MW)	$t_{i,u}$ (seconds)	g_i (MW/s)
1	90	80	1.2	15
2	32	160	1.3	8
3	10	100	1.3	2
4	200	5	1.4	25
5	62	0	1.4	15
6	25	10	1.5	5
7	56	40	1.5	6
8	81	50	1.6	12
9	8	84	1.7	1
10	27	18	2	6

Case 1 - Without Additional Frequency Limits

For the first case, following the solving methodology presented in Algorithm 5.1, the global minimum is found after optimising six regions. The global minimum is found in the last column of Table 5.6, which is labelled by the identifier (8,4,1). The identifier holds information about which region is being optimised, for more information consult Section B.1.2 about the meaning of each number. The preceding five columns are the local minimums of regions preceding the last region used to find the global minimum. It is noticed in the last row of Table 5.6 that the total cost incrementally improves after each step, as the total cost decreases.

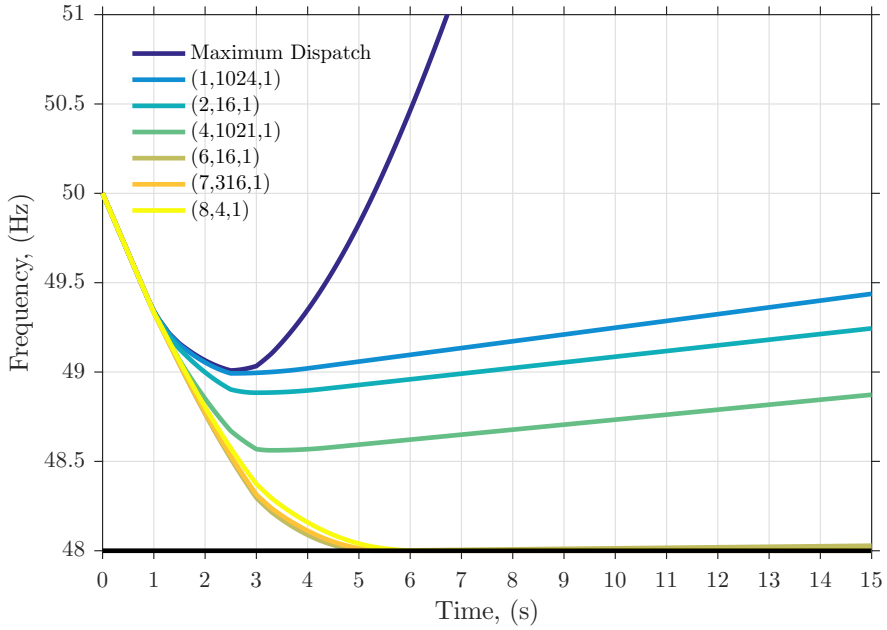


Figure 5.25 The frequency transient for the Second Example with only one frequency constraint.

It is not appropriate to analyse every step in detail, as that would be tedious, but this information can be found in the output from the MATLAB solver. There are a total of 7091 possible regions, not necessarily all of them have feasible points, but it is

not important to describe each region.

The frequency transient for each local minimum is shown in Figure 5.25. It is already recognisable that without the additional frequency limits the optimal solution will never create a dispatch that will result in the frequency returning to 50 Hz. This is confirmed by the total amount of reserve dispatched for the optimal solution, i.e. 400 MW equal to the risk of 400 MW, as shown in the last column of Table 5.6. Therefore the finally $df/dt = 0$ and the frequency transient remains constant at the limit.

Table 5.6 The local minimums after each optimisation to find the global minimum for the Second Example with only the initial frequency limit. The triple is an identifier for the region, the meaning can be found in Appendix B.

Step	1	2	3	4	5	6
Triple	(1,1024,1)	(2,16,1)	(4,1021,1)	(6,16,1)	(7,316,1)	(8,4,1)
IL Offer	IL Dispatch (MW)					
1	3.9	0	0	0	0	0
2	38	38	0	0	0	0
3	16	16	16	16	16	16
4	57	57	57	33.77	39.32	57
5	42	12	0	0	0	0
6	63	63	11.9	0	0	0
7	29	29	29	0	0	0
8	50	50	50	0	0	0
9	0	0	75	75	75	53.70
10	0	0	0	18	18	18
SR Offer	SR Dispatch (MW)					
1	19.5	27	31.5	0	0	0
2	9.6	0	0	0	0	0
3	2.4	3.4	4	0	0	0
4	27.5	40	47.5	97.5	103.33	121.33
5	39.25	43	45.25	60.25	62	62
6	5	7.5	9	19	20.17	21.96
7	6	9	10.8	22.8	24.2	13.34
8	10.8	16.8	20.4	39.63	20.78	16.8
9	0.8	1.3	1.6	0	0	0
10	3	6	7.8	19.8	21.2	19.88
	Total Dispatched Reserve (MW)					
	422.75	419	416.75	401.75	400	400
	Total Cost ($)					
	34,323	29,545	21,676	10,232	9,773	9,308

Case 2 - With Additional Frequency Limits

For the second case, additional frequency limits are added as shown in Table 5.7. Although the frequency limits do not express any particular restriction that might be expressed in electricity codes, they are chosen to reflect a reasonable rebound in the frequency transient. No analysis was undertaken to calculate the best time points or frequency levels of the limits, but they are used to illustrated the need for this analysis in practical implementations. This also applies to the offers, they do not reflect any real world situation, but are designed to illustrate the solver.

Table 5.7 The frequency limits added to the Second Example.

Time, $t_{j,l}$, (seconds)	Frequency (Hz)	Frequency (pu)
8	48.75	-0.025
9	49.35	-0.013
10.5	49.6	-0.008
12	49.8	-0.004

The results for the second case are shown in Table 5.8 and the corresponding frequency transients in Figure 5.26. Five regions are solved, resulting in five local minimums, the last of which is the optimal solution as seen in the last column of Table 5.8.

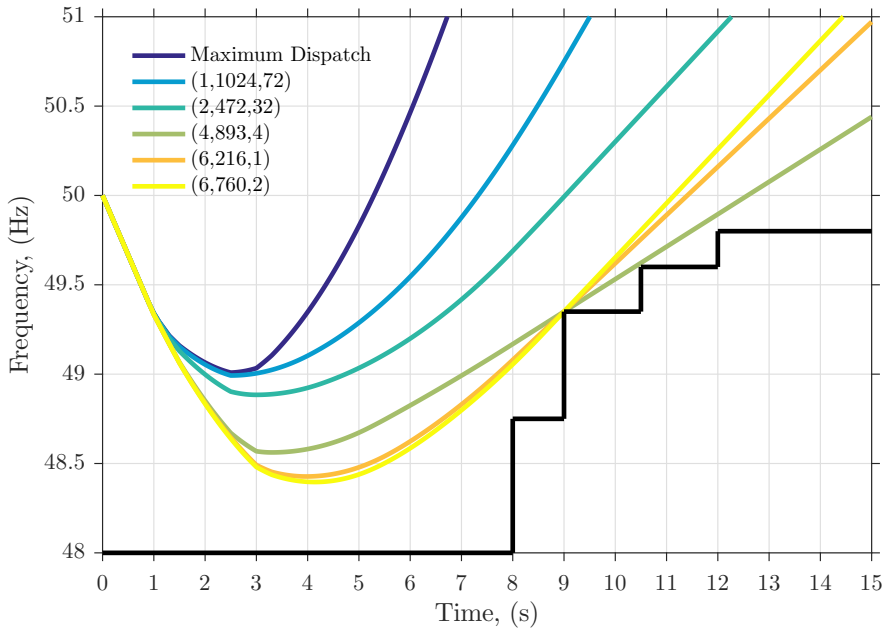


Figure 5.26 The frequency transient for the Second Example with the addition of more frequency constraints.

There is a clear difference between the two cases, with only one constraint the frequency does not rebound, and with additional constraints the frequency does, as shown by the yellow line in Figure 5.26. The total amount of reserve dispatched confirms this, for the first case 400 MW is dispatched to just cover the risk, but not enough for

Table 5.8 Results for the Second Example with the addition of more frequency limits.

Step	1	2	3	4	5
Triple	(1,1024,72)	(2,472,32)	(4,893,4)	(6,216,1)	(6,760,2)
IL Offer	IL Dispatch (MW)				
1	3.9	0	0	0	0
2	38	38	0	0	0
3	16	16	16	16	16
4	57	57	57	57	57
5	42	12	0	0	0
6	63	63	11.9	0	0
7	29	29	29	29	25.83
8	50	50	50	0	0
9	0	0	75	75	75
10	0	0	0	18	18
SR Offer	SR Dispatch (MW)				
1	19.5	27	31.5	31.91	25.16
2	9.6	0	0	0	0
3	2.4	3.4	4	1.55	0.09
4	190	165	94.78	165	180.43
5	39.25	43	62	62	62
6	5	7.5	9	12.35	25
7	54	45	39	29.96	26.63
8	76.8	16.8	20.4	51.19	42.88
9	7.3	6.3	1.6	1.38	0.87
10	3	6	7.8	11.81	27
	Total Dispatched Reserve (MW)				
	705.75	585	508.98	562.15	581.90
	Total Cost ($)				
	40,901	32,030	23,040	18,399	17,289

derivative of the frequency to ever be positive. For the second case, the total reserve dispatched is 581.90 MW, as shown on the second to last row and last column of Table 5.8. This is significantly more than the first case, but not too much that frequency rebounds at the same rate at which it dropped, which would occur if 800 MW were required.

The rate at which the frequency rebounds, or the amount of reserve needed over and above the risk requirement, is partly influenced by the timing of the frequency limit function. To understand the impact of timing on these constraints have, the marginal effect is calculated if the frequency limits are delayed from the second case. Delaying by 0.1 seconds, the total reserve requirement counter-intuitively increases by 0.38 MW. However, the cost reduces by $238, as costly reserve from IL offer 7 can be replaced by SR offers, mainly from SR offer 4, but it needs slightly more reserve from the SR offers to replace the IL offer and still satisfy the frequency constraints, hence the increase in total reserve required. The methodology for finding the marginal value is given in

Appendix C.

The timing of the frequency limits has a significant influence on the dispatch and cost, the choice is not trivial. Any market implementation is going to require a clear decision methodology, but there is no explicit code or policy that would determine this, as all equipment is required to remain connected at these frequencies. A possible direction for determining these constraints is to align them with historic frequency transients, so that they reflect capabilities of the existing grid. However this may be too restrictive in light of new technologies.

Marginal Reserve Provider

There does not appear to be a marginal generator, as there are six SR offers and one IL offer not on their limits, as seen in the last column of Table 5.8 for the second case. There is a consistency to the dispatch, the most expensive reserve providers are not dispatched and the least expensive are, as is expected in an optimisation that is not overly constrained. Also for every 1 MW the risk rise, IL offer 7 increases 1.2 MW.

The reason why there are six marginal SR offers can be attributed to the single binding frequency constraint, and a concept of effective area. The value of a single MW of reserve in satisfying the frequency constraint depends on when that MW is provided and how close it is to the time of the frequency limit. Define τ_i as the time that MW is provided, then the effectiveness of that MW is $t_{j,l} - \tau_i$, as the longer that MW is before the limit the longer that MW is integrated and can increase the frequency. Hence the concept of area, not that reserve is required, but reserve over time. That is equivalent to the area bound by reserve imbalance curve from the right hand side of Eq. 5.4. Now each MW of reserve has a different cost, therefore the price of that effective area is:

$$\eta_i(\tau_i) = \frac{c_{i,u}}{t_{j,l} - \tau_i} \tag{5.178}$$

The optimisation process finds the optimal effective area price and dispatches all reserve below that price. This price is found as the Lagrangian of the binding frequency constraint, \$13.07/MW s for the second case. This price refers to effective area price of IL offer 7, 13.07 = 98/(9 - 1.5). To show that the SR offers are dispatched up to this price, consider SR offer 4, it is dispatched at 180.43 MW, it would then have to stop ramping at 8.62 seconds, the effective area price is $5/(9 - 8.62) = 13.07$.

So in this situation of one binding frequency constraint there is a concept of a single marginal reserve provider, but its form is different to the normal formulation. Instead of a merit-order ranking each MW of reserve by price in \$/MW, each MW of reserve is ranked by effective area price in \$/MW s. The marginal provider is set by the area requirement for the frequency constraint, as opposed to the reserve covering the risk. This difference is shown in Figure 5.27. There are still difficulties with the concept of a marginal generator when more than one frequency constraint are binding and especially

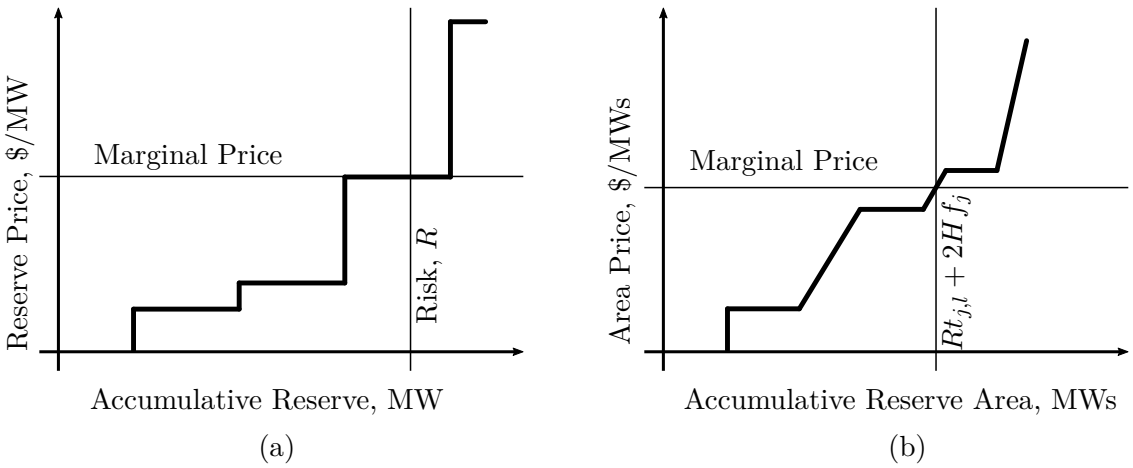


Figure 5.27 The difference between a typical merit-order for a standard reserves market, (a), compared against the merit-order for the new formulation, (b).

when it is the minimum frequency constraint.

The concept of effective area also explains the marginal increase in total reserve with rising risk. The added risk increases the area requirement, the risk's effectiveness is 9 seconds compared to IL offer 7's, 7.5 seconds. Therefore the reserve has to increase in proportion, $9/7.5 = 1.2$.

Solver Performance

The performance of the solver is measured by how many regions are optimised, as optimising each region can be roughly equated with optimising a single real-time co-optimised energy and reserves market that uses a LP formulation, if this new formulation were applied to that same real-time market. For the second example, attention is placed on comparing the differences in performance when including additional frequency limits. In Table 5.9 key metrics of each case are listed. Eq. 5.98 provides an estimate of the total number of possible regions, these estimates are shown on the first row of Table 5.9. Clearly, including additional frequency limits should increase the total number of possible regions, and this is seen in the second row of Table 5.9 where actual number is given. Adding four frequency limits has multiplied the number of possible regions by 15. It is also noticed that the actual number is significantly less than the estimate.

A distinction is made between 'possible' and 'feasible' number of regions. 'Possible'

Table 5.9 Performance of solving each case of the second example

	Without Freq. Limits	With Freq. Limits
Est. No. of Possible Regions	22,528	1,572,120,576
No. of Possible Regions	7,091	106,437
No. of Feasible Region	7,071	91,713
Est. No. of Solved Regions	39	43
No. of Solved Regions	6	5

refers to how many regions above the reserve requirement plane can be made. However not all of these regions will have points that satisfy the frequency constraints, therefore another total called 'feasible' is used to count these.

In Eq. 5.128 an estimate of how many regions are required to solve a problem is given. The estimates for the second example are shown on the fourth row of Table 5.9. It is apparent that including additional frequency limits has minimal effect in increasing this estimate. When it comes to solving each case, it is apparent that including additional frequency limits has no impact, as the number of regions solved reduces from six to five. It is not fair to make a generalisation from one example, but it has been shown that including additional frequency limits will not necessarily degrade solver performance.

5.4.3 Third Example - Optimal IL Region

So far from the first two examples the optimal region has been of SR type, i.e. the global minimum has been found in a SR type region. It is of interest to consider what a frequency transient would look like if the optimal region were of the IL type. To force this requirement, consider a problem with only IL offers, these are shown in Table 5.10. Inertia and risk remain at 15,000 MW s and 400 MW respectively, and the frequency constraints are repeated from Table 5.7 with the initial limit of 48 Hz.

Table 5.10 List of IL Offers for the Third Example.

	Quantity (MW)	Price ($/MW)	$t_{i,p}$ (seconds)
1	128	200	0.9
2	13	0	1
3	59	160	1.1
4	174	150	1.3
5	38	40	1.5
6	64	10	1.6
7	119	5	2
8	132	16	3

Table 5.11 Steps in the optimisation process to find the global minimum for the Third Example.

Step	1	2	3	4
Triple	(1,1,1)	(2,1,1)	(3,1,1)	(4,1,1)
IL Offer	IL Dispatch (MW)			
1	116	52	0	0
2	13	13	13	13
3	59	59	0	0
4	174	174	166	102.28
5	38	38	38	38
6	0	64	64	64
7	34.55	39.03	119	119
8	0	0	56.24	132
	Total Dispatched Reserve (MW)			
	434.55	439.03	456.24	468.28
	Total Cost ($)			
	60,433	48,295	28,555	20,209

When there are no SR offers there are fewer regions, there are only four possible regions, all of them are feasible, and all four are solved to find the optimal solution. The results are shown in Table 5.11 and Figure 5.28 with the optimal solution found in the last column of Table 5.11. The optimal solution avoids the expensive IL offers with a quick response, hence having to find the optimal solution in the last region where the time of the minimum frequency can be the latest. The main difference between solutions in IL and SR type regions is the frequency derivative at the time of the minimum frequency is discontinuous in the former and continuous in the latter. Discontinuity

occurs in New Zealand frequency transients in certain situations, this formulation can capture those events.

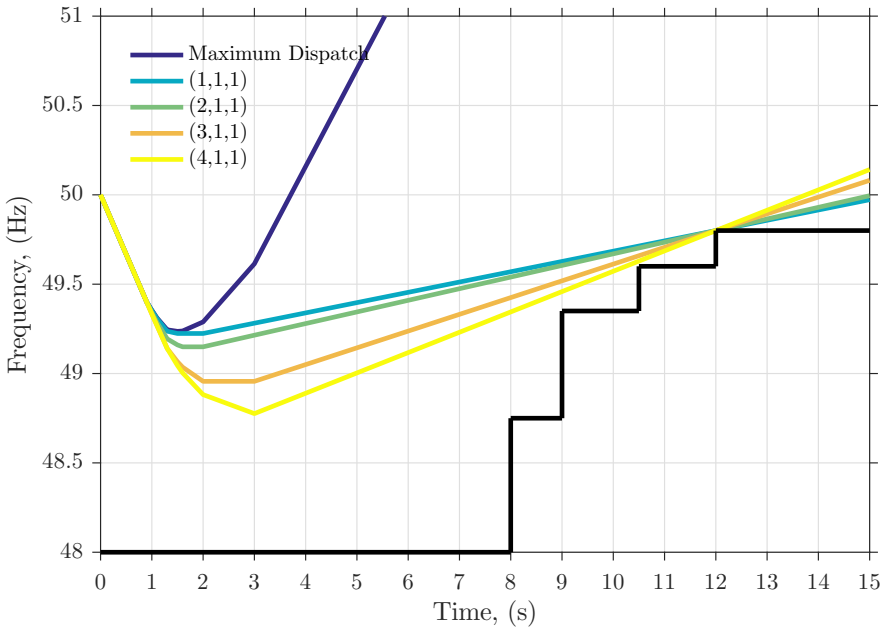


Figure 5.28 The frequency transient for the Third Example.

5.4.4 Fourth Example - Inertia and Risk Variation

This example is interested in how the dispatch changes with variations in inertia and risk. It is anticipated that with reduced inertia the amount of fast acting reserve required will increase. If the faster reserve is more expensive then the total cost will increase significantly. This example, for one set of offers, analyses the total cost for a range of inertia values to determine whether that relationship is true. Secondly, the performance of the solver is analysed in how many steps are required to solve the problem depending on risk and inertia.

A set of offers is presented in Tables 5.12 and 5.13. The offers do not reflect an actual situation, but are chosen for the purposes of illustration, like it is for the other examples. The offer prices are chosen to have a correlation between speed and price, with the fastest responses being the most expensive. The frequency limits of the last two examples is applied but delayed one second further, Table 5.14.

Table 5.12 List of IL offers for the Fourth Example

	Quantity (MW)	Price ($/MW)	$t_{i,p}$ (seconds)
1	68	400	0.9
2	16	300	1
3	54	200	1
4	152	160	1.2
5	23	120	1.8
6	89	80	2.5
7	48	0	3.5

Table 5.13 List of SR offers for Fourth Example

	Quantity (MW)	Price ($/MW)	$t_{i,u}$ (seconds)	g_i (MW/s)
1	71	150	0.6	16
2	26	130	0.8	4
3	67	110	1.1	5
4	62	90	1.2	12
5	165	70	1.6	18
6	47	50	2	6
7	33	30	2.2	3
8	28	10	3	6

Table 5.14 The frequency limits of the Fourth Example.

Time, $t_{j,l}$, (seconds)	Frequency (Hz)
9	48.75
10	49.35
11.5	49.6
13	49.8

Inertia Variations

This problem has standard inertia and risk parameters of 15,000 MW s and 400 MW respectively for the base problem. Inertia is varied from 6,500 MW s to 65,000 MW s: the changes in total cost for the optimal solutions are shown in Figure 5.29, frequency transients are shown for several different inertia values in Figure 5.30, and the optimal dispatch for some inertia values is shown in each column of Table 5.15. The minimum feasible value for inertia is 6,433 MW s, there is no upper limit.

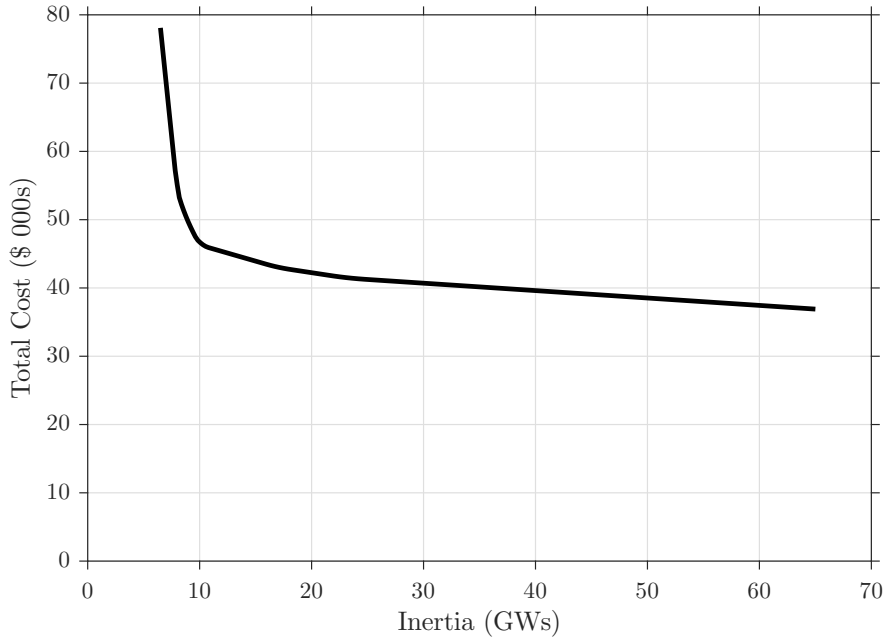


Figure 5.29 Differences in total cost for variations in inertia.

It has been proved correct that with decreasing inertia the total cost does increase significantly, for these offers this occurs around 10,000 MW s. Looking at the frequency transients of Figure 5.30, the curves align with expectations, at higher inertia the frequency transient does not reach the minimum of 48 Hz, but rather the last frequency limit at 13 seconds is the binding constraint. When the inertia decreases, the binding frequency limit swaps to an early time and eventually the 48 Hz limit becomes binding. At 6,500 MW s the frequency transient remains flat at the 48 Hz limit for about a second, as the system waits for the less expensive reserve to return it back to 50 Hz. Also at the lowest inertia, and because of the amount time available, the amount of reserve required to return the frequency back to 50 Hz is less. Therefore the total reserve dispatched is less, and can be seen by comparing values on the second to last row of Table 5.15.

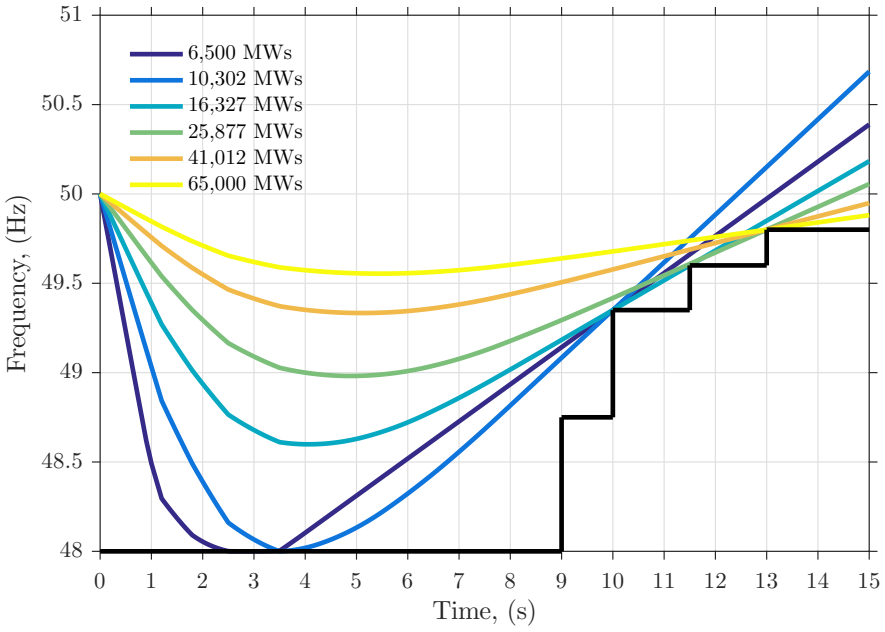


Figure 5.30 Frequency transients for different inertia values.

Table 5.15 Results for Fourth Example for variations in Inertia.

Offer	Inertia (MWs)					
	6,500	10,302	16,327	25,877	41,012	65,000
IL Offer	IL Dispatch (MW)					
1	65.49	0	0	0	0	0
2	16	0	0	0	0	0
3	54	0	0	0	0	0
4	152	134.26	95.23	49.59	39.33	23.06
5	23	23	23	23	23	23
6	15.01	89	89	89	89	89
7	48	48	48	48	48	48
SR Offer	SR Dispatch (MW)					
1	24.8	15.99	18.67	21.4	21.4	21.4
2	5.8	6.39	8.40	10.45	10.45	10.45
3	6.25	10.47	14.67	18.94	18.94	18.94
4	15	37.03	47.60	61.95	61.95	61.95
5	17.11	71.21	84.60	112.28	112.28	112.28
6	3.9	28.95	32.60	43.88	43.88	43.88
7	1.65	17.69	19.10	25.76	25.76	25.76
8	6	28	28	28	28	28
	Total Dispatched Reserve (MW)					
	454	510.00	508.85	532.24	521.98	505.71
	Total Cost ($)					
	78,090	46,319	43,310	41,147	39,505	36,903

Risk Variations

Retaining a constant inertia of 15,000 MW s, risk is varied from 200 to 600 MW. Total cost of the optimal solutions to variations in risk are shown in Figure 5.31. For selected values of R, the optimal frequency transients are shown in Figure 5.32, and the optimal dispatch of reserve is shown in each column of Table 5.16 for those same values of R. The maximum feasible risk is 626.76 MW, the lower limit is a risk of zero, but practically it should be above this.

It is clear that variations on risk has a different impact than variations on inertia to cost and the optimal reserve dispatch. Variations in risk provide a more consistent change, as total cost increases linearly in Figure 5.31. While the changes in cost for variations in inertia are mostly flat for wide range of values, risk is the opposite. Therefore risk should always be a variable in a co-optimised market. The results also have implications for questions about creating inertia markets: if it is uncommon for expensive fast resources to be dispatched, due to the minimum frequency constraint being mostly unconstrained, then an inertia market is unnecessary. However, further analysis is required to substantiate these claims, as these generalisation are made from a single example and without sufficient definition of terms.

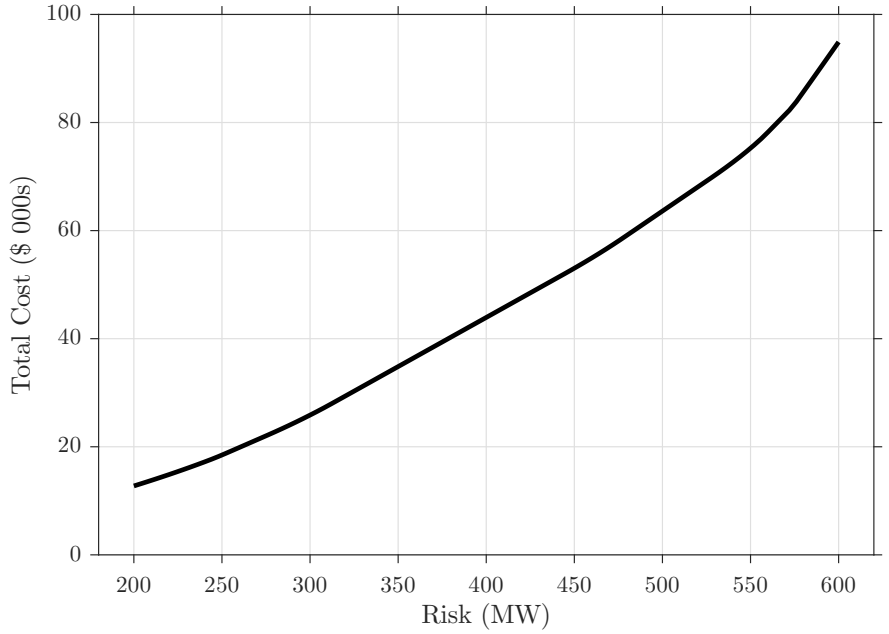


Figure 5.31 Differences in total cost for variations in risk.

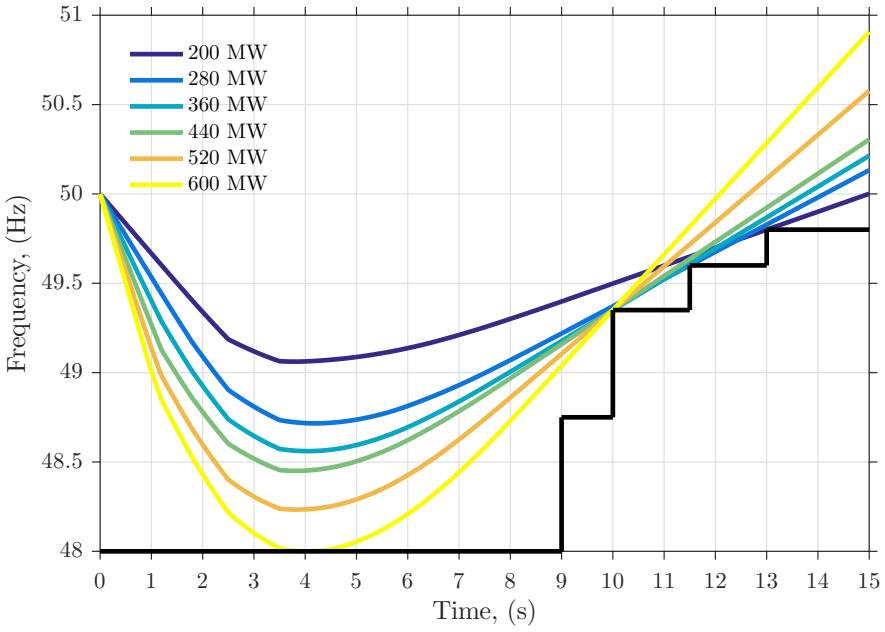


Figure 5.32 Optimal frequency transients for different risk values.

Table 5.16 Results for the Fourth Example with variations in Risk.

Offer	Risk (MW)					
	200	280	360	440	520	600
IL Offer	IL Dispatch (MW)					
1	0	0	0	0	0	33.43
2	0	0	0	0	0	16
3	0	0	0	0	46.26	54
4	0	0	56.77	147.68	152	152
5	0	23	23	23	23	23
6	89	89	89	89	89	89
7	48	48	48	48	48	48
SR Offer	SR Dispatch (MW)					
1	0	0	18.4	18.4	42.4	69.38
2	0	2.00	8.2	8.2	13.4	19.25
3	0	8.85	14.25	14.25	19.75	25.93
4	3.18	38.87	46.2	46.2	57	62
5	43.71	79.35	81.9	81.9	94.5	108.66
6	27.55	33.46	31.5	31.5	34.5	37.87
7	20.86	20.84	18.45	18.45	19.35	20.36
8	28	28	28	28	28	28
	Total Dispatched Reserve (MW)					
	260.30	371.38	463.67	554.58	667.16	786.88
	Total Cost ($)					
	12,748	22,745	36,656	51,201	68,057	94,903

Solver Performance

Performance of the solver is analysed by counting the number of regions solved to find the global minimum. In Figure 5.33 for variations in inertia, the total number of feasible regions is compared to the total number of solved regions, and likewise in Figure 5.34 for variations in risk. From visual inspection of the two Figures, the total solved regions shows good correlation with the total number of feasible regions, which agrees with the idea that if the solution space is bigger then it will require more steps to traverse across it. However this correlation is lost when the problem becomes more constrained, as the number of feasible regions decrease, when inertia is low and when risk is high. It is unknown why this correlation is lost. The total number of steps also shows an unpredictability, as it fluctuates up and down without any obvious reason. It is difficult to accurately predict the number of solves based on simplified information, such as the number of offers and frequency limits. Even without the consideration of risk the prediction can be significantly wrong, as for this example, it fluctuates between five to thirteen regions.

The estimate of Eq. 5.128 anticipated a maximum number of solved regions to be 33 for this fourth example. However, the maximum used of all the simulations, i.e. all inertia and risk variations, was 13. This is significantly better than the estimate. Therefore it can be antipated that the estimate of Eq. 5.128 will over estimate, and the actual value will be significantly more practical.

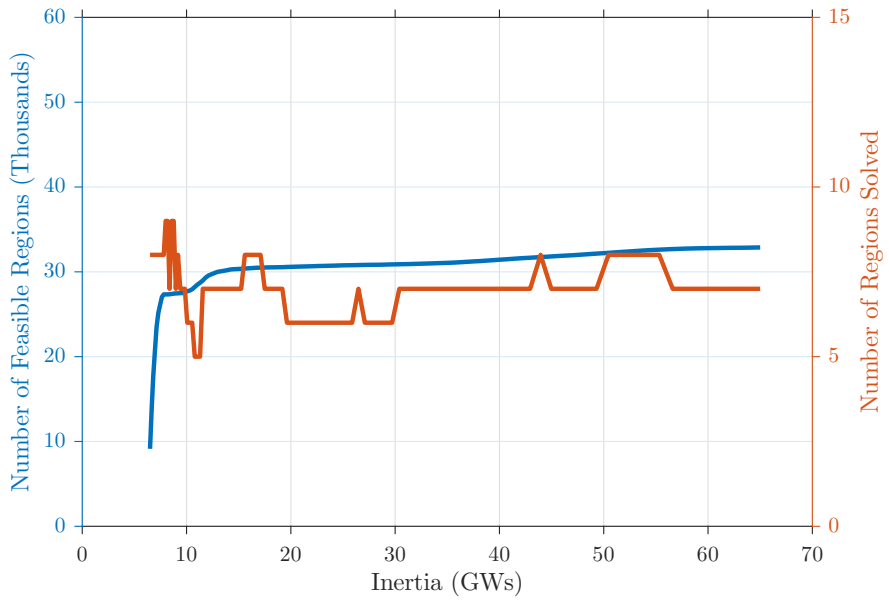


Figure 5.33 The number of feasible regions, and the number of regions solved in order to find the global minimum, for variations in inertia.

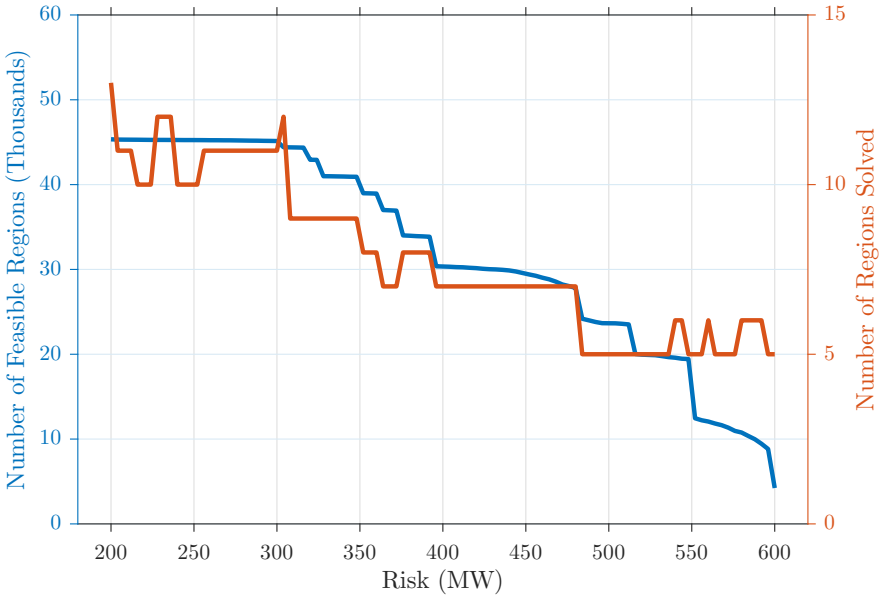


Figure 5.34 The number of feasible regions, and the number of regions solved in order to find the global minimum, for variations in risk.

5.4.5 Conclusion

Through these four examples, the key features of this formulation have been presented. Four main conclusions are made. (1) Additional frequency limits, after the initial one, are critical to ensuring enough reserve is dispatched for the frequency transient to return back to 50 Hz. The cost of adding these extra limits to the solve time is minimal. (2) The optimal solution is not likely to conform to a dispatch based on the standard merit-order curve, i.e. by ranking reserve prices. The optimal solution can be derived, in singular bound frequency constraint situations, by a merit order-curve that ranks reserve area prices. (3) Optimising dispatch to minimise the largest credible contingency, i.e. risk, is more important than optimising for the total rotational energy on the power system to be optimised, i.e. inertia. (4) This new reserves optimisation formulation can be practically applied in real-time energy and reserves markets, as solve times for the problems shown are practical. This is provided that further research is done, which is discussed in the next chapter.

Chapter 6

FUTURE DEVELOPMENT

This thesis has presented a means of optimising reserve providers with varying response rates. This chapter focuses on directions that should be taken to develop this formulation into a practical tool for any electricity industry. Not just as an addition to an electricity market, but for investment decisions as well. Future development is focused on two key aspects: what is necessary to solve the co-optimised energy and reserves problem, and how can reserve offers with finite duration be incorporated, particularly responses from wind turbines. Lastly general recommendations are given.

6.1 CO-OPTIMISED ENERGY AND RESERVE

The addition of energy constraints into the formulation is not a major difficulty in principle, as energy constraints from MILP formulations are linear. It is just a matter of adding the extra equations to each region. However, there are three difficulties that complicate the solving methodology:

1. The solving methodology requires a known feasible point to find the first region to solve. The current methodology chooses the point where all reserve is dispatched at its maximum output. For a co-optimised formulation, where generation capacity is shared between energy and reserves, this maximum dispatch point may not always be feasible. This raises two important questions for any problem: does a solution exist, and where is a good starting point?

2. In the New Zealand power system, the largest credible contingency (which is also called the risk) comes from the HVDC link between the North and South Islands, or from the largest generator. In the latter case, reserve that a generator provides cannot be used to cover its own risk, as this is an impossibility. So far risk has been considered independent from any reserve provider, and the largest credible contingency does not also provide reserve, but this is unsatisfactory for a practical implementation. To fix this problem a potential frequency transient is created for each potential contingency, and the reserve from that risk does not

feature in the equations. Hence for each potential contingency there is a set of reserve requirement equations, frequency constraints, and region boundaries. The problem still remains convex, as the intersection of finitely many convex sets is also convex itself, and it is still possible to solve, but complexity has jumped significantly. The feasible space has divided into smaller regions, requiring a larger set of transition rules to direct it towards the optimum. Considering how long the rule list is in Appendix B, developing a significantly longer list is not encouraged. A better method is to have a general approach to choosing the next region based on the local geometric properties.

3. Many real-time electricity markets employee mixed integer variables into their formulation. For the new reserves optimisation to be incorporated into these formulations, the solving methodology has to be developed to include mixed integer variables.

The two main issues are determining the feasibility of a problem, and developing a general solving methodology. These two problems also arise when including reserve of finite duration, the next topic.

6.2 NEGATIVE RESERVE AND WIND RESERVE

Two types of reserve have been optimised in this thesis, reserve that responds instantaneous (Interruptible Load, IL) and reserve that responds over a finite duration (Spinning Reserve, SR). Although the names given to these in the thesis suggest a limit to the types of reserve they can be model, they are just practical descriptions for those familiar with the New Zealand reserves market. These two types of reserve can model a whole range of responses, just with the right choice of the parameters, $t_{i,p}$, $t_{i,u}$, and g_i. More difficult reserve types can be produced from multiple offers. This could possibly include different forms of energy storage. However there are potential sources of reserve that cannot be accurately modelled by these two types. Synthetic Inertia from wind turbines, loads that can only disconnect for a finite duration, and reserve providers that can be over-rated for a short duration, are reserve responses that come to mind. This section envisages that it is possible to optimise these forms of reserve also, i.e. a convex space can be formed, and a global minimum can be found by shifting from one region to another, but with more complexity. Convexity is not proved for these reserve types, but it is hoped future research will show this.

6.2.1 Negative offers

IL offers are an instantaneous positive real power output. To form other reserve types that can only sustain output for a short duration, it is necessary to allow for negative instantaneous real power output. This is referred to as Negative Reserve (NR), which is illustrated in Figure 6.1. NR is initiated at time, $t_{i,m}$, and reaches a negative output with absolute magnitude of m_i MW. The next section demonstrates how NR with IL and SR offers can form more complicated reserve providers, but now consider what equations are required to keep the frequency transient above the frequency limit function. The problem for the most part remains the same: there is the reserve requirement constraint, frequency constraints at the time of transitions, $t_{j,l}$, and offer limits. The point of difference is the minimum frequency constraint, not that it is removed, but that there could be multiple minimum frequencies, as shown in Figure 6.2. In Figure 6.2 with the addition of one NR offer that responds at $t_{1,m}$, there is the possibility of having two minimum frequencies.

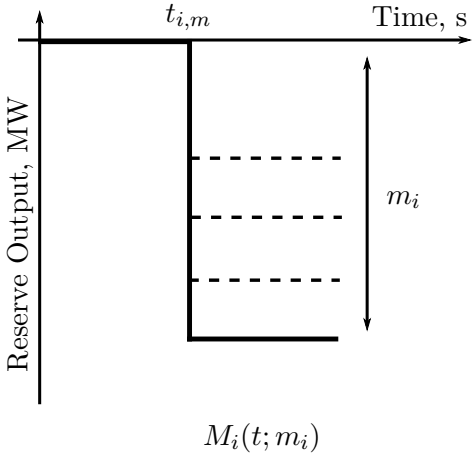


Figure 6.1 Negative Reserve Offers

Each time a negative reserve offer occurs, there is a possibility the frequency transient could revert to falling again, and another minimum frequency constraint is required for the system to remain stable. Instead of one minimum frequency constraint, there is a potential extra one for each unique time there is a NR offer. In the original formulation, the types of regions are divided into two, IL type region and SR type region, but with NR there is a multiplicity. For example, within a single problem, in addition to the standard regions there is also the possibility of having SR-SR type, IL-SR type, and other type regions. The first minimum frequency could be of SR type and the second one is of the SR type also, thereby creating the SR-SR type region. In addition, an extra internal boundary between regions is found whenever the total reserve at $t_{i,m}$ is equal to zero.

It is anticipated that this problem will remain convex for two reasons. It has already been shown that the space defined by the frequency limits is convex in Section 5.2.2.

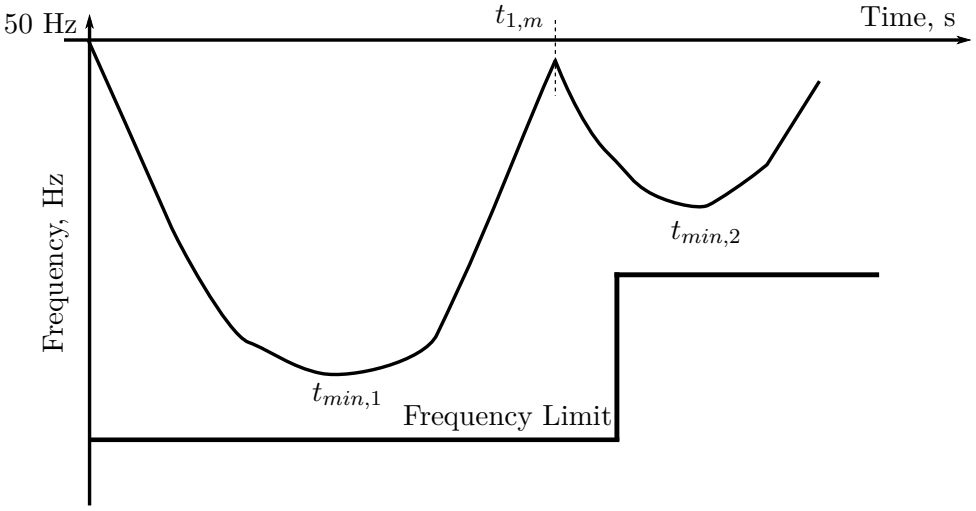


Figure 6.2 Possible frequency transient with one Negative Reserve offer.

It did not matter whether the IL offers where positive or negative. The second reason is for the minimum frequency constraint, since the structure of the problem has not changed significantly, the same principles can be applied to the multiplicity of minimum frequencies. In the proof of convexity, when showing all points along the line between $\mathbf{v}_1$ and $\mathbf{v}_2$ are also elements of the same space, as λ transitions from $\lambda = 0$ to $\lambda = 1$ in the parametrisation of that line, the time of the minimum frequencies will bifurcate and combine. This is what separates it from the current proof.

Since the opposite response to IL offers is being considered, what about the opposite response to SR offers? Then the number of possible reserve types that can be modelled can increase even further. These can be better approximations of wind reserve. However it is anticipated that including this form of reserve will make the problem non-convex. For the same reason that kept SR offers convex in the frequency limits, the opposite will make it concave. The overall problem may be convex once the minimum frequency constraints have been added, but this is not likely given the frequency limits. Therefore the opposite to SR offers is not recommended straight away. In certain situations under extra constraints it may be possible. Nevertheless it eventually needs to be shown if it is convex or not.

6.2.2 Wind Reserve Offers

The easiest way to create a wind reserve offer is from two IL offers and one NR offer, as shown in Figure 6.3, with the addition of two equality constraints determining the relationship between the offers. This is to emulate an inertial response when the wind turbine is not curtailing generation. The wind turbine when operating at its maximum power point, can increase power output further while slowing down the speed of the blades, drawing extra rotational energy from the blades to provide reserve. This cannot be sustained for long as the rotor will stall, but more importantly before this happens

the operating condition will deviate too far from the maximum. A recovery period is required to restore the system, where output is reduced below pre-contingency levels, bringing the speed back up to optimal.

The initial reserve output is modelled by an IL offer, p_1 at $t_{1,p}$ as shown in Figure 6.3, which is sustained until the start of the recovery period. A reduction in power output is made at $t_{1,m}$ with a absolute magnitude of m_1 to start the recovery period. A relationship is required between m_1 and p_1, so that the total overall energy lost from deviating from the maximum power point is accounted for. Therefore introduce a scaling parameter, $e_1 < 1$, that approximates a greater amount of energy required for recovery than for reserve provision:

$$\begin{aligned} \textit{reserve provision energy} &= e_1 \times \textit{ recovery energy} \\ p_i(t_{1,m} - t_{1,p}) &= e_1(m_1 - p_1)(t_{2,p} - t_{1,m}) \end{aligned} \tag{6.1}$$

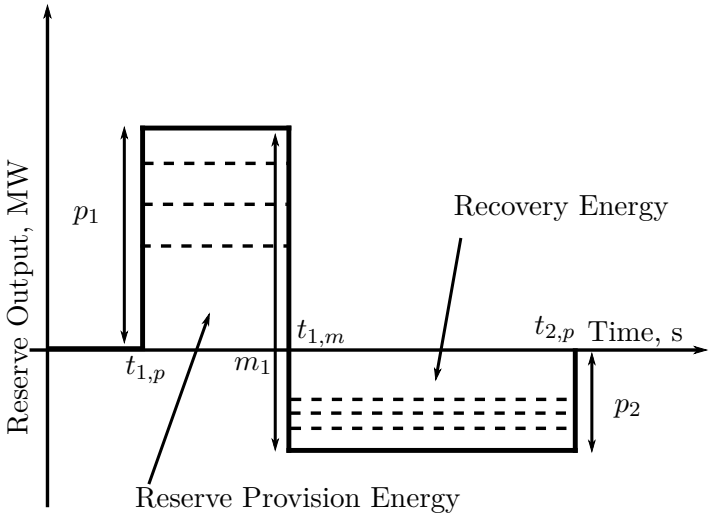


Figure 6.3 Wind reserve offer to emulate an inertial response from a wind turbine.

The final IL offer, i.e. p_2 at $t_{2,p}$, is required to return the total power output back to pre-contingency level: $p_2 = m_1 - p_1$. This proposed wind reserve offer represents only one possible way of creating an offer. Other combinations of offers could provide a different response to best model other reserve types.

6.2.3 Feasibility and Solving

Two main issues arise with the introduction of wind reserve offers, just as it did with co-optimising energy and reserves: feasibility and solving methodology. The presence of a region where all positive reserve is dispatched and all negative reserve is avoided, i.e. the starting point of the solving methodology is no longer available due to the relationships like Eq. 6.1. It is not known with wind reserve whether it contributes positively or negatively to finding a feasible solution. Secondly with the introduction of

negative reserve offers, the number of regions increases, and the transition rules of the solving methodology are more complex.

It is recommended that an algorithm to determine feasibility be developed. There should be a significant amount of research done in the Optimisation and Operations fields already, which will likely involve a better analysis of the dual space, to solve this problem. If a new approach is required, then a method of jumping from one internal vertex to another and calculating the feasibility of each point could be possible.

To simplify the solving methodology, a general geometric approach would be best. The problem of transitioning from region to region is a matter of finding the internal hyper planes that intersect the local minimum, and then finding which side of the hyper planes the global minimum could be found in. An even better approach would be to combine the two hierarchies of the solving methodology: the Newton steps in solving each individual region, and the region transition steps. Allowing the Newton steps to jump across internal boundaries would give the best computational performance, and simplify the problem. However, the dual space needs better analysis, and the minimum frequency constraints will add difficulties due to their discontinuities.

6.3 GENERAL RECOMMENDATIONS

General recommendations are listed as follows:

1. A pricing methodology is needed for a market implementation. The creation of one has been avoided in this thesis. A potential approach could be to start with the marginal cost of risk from the constraining generator, and then scale the reserve price for each provider based on the performance characteristics of each. This should incentivise faster response over slower, so that the best investments are made to reduce overall electricity costs. An approach like this will require clear justification, as competition between providers will want to manipulate the scaling in their favour.

2. The computational expense of this formulation needs to be understood for market implementations. So far the algorithm performance has been measured in how many equivalent SPD type solves (MILP) are required. A maximum estimate is given in Eq. 5.128, but this significantly overestimates the actual number required. A formal estimate is required.

3. Frequency transients are found by solving Eq. 5.4. In this equation there is no feedback to account for load dynamics. Although this is not an issue as load dynamics can be approximated by IL and SR offers (this should be analysed too), it would be useful to know if having feedback whether the problem still remains convex, or what conditions on the feedback parameter will make it convex. If

found to be so, then what about higher order differential equations? For example, incorporating reserve from the South Island into the North, by having both the North and South Island frequencies as sate variables, plus the HVDC transfer, each of them having limits. Therefore an accurate optimisation can take place for contingencies that obtain support from the other Island. This area of research leads heavily into optimal control, and solutions to these questions may be found there. These problems are of low importance as contingencies that do not separate the two islands are sufficiently well managed already, and the impact of Variable Renewable Energy (VRE) in the future is only going to reduce this problem, Section 3.7.

Chapter 7

CONCLUSION

This research has developed an optimisation process to find the optimal dispatch of primary reserve. The motivation, in the New Zealand context, came from an anticipated increase in peak Instantaneous Reserve (IR) requirement with higher penetrations of Variable Renewable Energy (VRE), resulting in lower rotational energy and higher credible contingencies, which increases demand for reserve capacity with sufficient speed. Although the average reserve requirement is expected to decrease, higher HVDC transfers north would increase risk and reserve requirements for short periods. The current formulation for IR optimisation suffers from inefficiency under these situations, by RMT demanding greater amounts of FIR to ensure sufficient speed of response. The result is some capacity is underutilised, which may be valuable for energy production.

Inefficiency in the current formulation came from translating dynamic requirements in RMT to the linear constraints of SPD. The formulation developed here removes this inefficiency by transferring some of the dynamic requirements to the optimisation formulation. The result has not retained linearity, but has satisfied convexity thereby keeping the important features that are desirable in LP problems, such as uniqueness of solution and practical solve times. However discontinuities in constraints increase computational load, as equations change depending on region. It has been shown that this extra computational load is reasonable within the 5 minute dispatch window, but further work is required to match performance speed against current MILP solvers.

The new formulation introduces the swing equation into the optimisation, connecting the reserve offers and their transient features to the frequency limits that determine the dynamic requirements for reserve. The concept of adding dynamic equations to optimisation is not new as the field of Optimal Control focuses on these problems. However the way it is done here is new, which avoids the complexities of Continuous Optimal Control, and the computational burden of Discrete Optimal Control with its expansion of variables, which Miller [2014] and Sokoler et al. [2016] achieve. The new approach simplifies reserve responses over time by only allowing variations in two response categories: IL with its step increase in power output, and SR with its ramped power output. These responses can reflect many different categories of reserve provider

through the choice of $t_{i,p}$, $t_{i,u}$, and g_i, e.g. a battery system, since it responds very quickly, can be approximated by an IL response with $t_{i,p}$ defined by how fast it is. Since any number of these reserve responses can be added to a problem, competition is maintained without having to separate providers into different reserve classes.

Although the impetus for this research was found in the New Zealand context, there is no reason why it cannot be applied to overseas power systems as well. For Australia, Ireland, and the United Kingdom their primary reserve products can be combined with their potential enhanced reserves, and even synthetic inertia could be added if Negative Reserve (NR) is a possible option. For Ireland especially, this could remove the System Non-Synchronous Penetration (SNSP) limit which is likely to unnecessarily curtail wind generation.

Appendix A

PROOF OF MAIN RESULTS FOR COMPLEXITY

In Section 5.2.3, the convexity of the feasible solution space is proven. The proof is dependent on results that are shown here in this Appendix: the nature of the time of the minimum frequency, $t_{min}(\lambda)$ (often called the minimum time), and the nature of the area function, $A(\lambda)$, which determines the minimum frequency. A framework is created for the proofs, so that every situation can be rigorously analysed as λ progresses from 0 to 1. The notation in this Appendix follows Section 5.2.3, and a list of new symbols is provided just below.

A.1 NOTATION

Parameters

Ω	A set of all time period boundaries
$\tau_{j,s}$	Start time of the j region
$\tau_{j,e}$	Finish time of the j region
$\Delta_{i,j}$	Reserve offered by i in region j
$\theta_{i,j,s}$	Total amount of reserve available from offer i by time $\tau_{j,s}$
$\theta_{i,j,e}$	Total amount of reserve available from offer i by time $\tau_{j,e}$
$B_{l,0}$, $B_{l,1}$, and C_l	relate the progress of γ_j with λ, look to Eq. A.17 and A.45 for a proper definition.

Counts and Totals

$Nreg$	Total number of regions
$Noff$	Total number of offers

Functions

$OFF(i)$	Type of offer from i
$REG(i)$	Type of region for j

Variables

$\mu_{i,j}$	The proportion of $\Delta_{i,j}$ dispatched for the i offer
γ_j and γ	Position in regions where reserve covers risk
S and E	Variables in the domain of possible regions, S referring to the left most region and E to the right most region

A.2 SEPARATING RESERVE INTO BLOCKS

The minimum time changes depending on the quantity of reserve that finishes ramping before t_{min}. The goal is to mathematically describe this relationship. If the amount of reserve before t_{min} is decreasing, then the minimum time has to increase to compensate, and vice versa. For each λ there is a set of critical offers that are still ramping up at t_{min}. The rate at which t_{min} varies is dependent on the slope of those critical offers. As λ increases, eventually there are transitions where the critical offers change. It is beneficial to separate reserve into time periods it is ramping and the ones it is not.

Firstly a simplification is made, if an offer is dispatched to zero for both points, $\mathbf{v}_{i,1} = 0$ and $\mathbf{v}_{i,2} = 0$, then that offer is no longer considered as it has no influence on either t_{min} or $A(\lambda)$. Also following Section 5.2.3, all IL offers with the same initiation time are combined into a single offer.

The time periods (or regions) are separated by the time SR starts and stops ramping, and the time IL is initiated. The maximum time SR can stop ramping is $t_{i,e}^{max} = t_{i,u} + u_i^{max}/g_i$. A set of all the region boundaries is:

$$\Omega = \{t_{i,p}\} \cup \{t_{i,u}\} \cup \{t_{i,e}^{max}\} \tag{A.1}$$

Each region, indexed by j, is created with a start time $\tau_{j,s} \in \Omega$, and a finish time $\tau_{j,e} \in \Omega$. If there is an IL time $t_{i,p}$ then there is an extra region where both the start and finish times are the same and equal to $t_{i,p}$. Each region is ordered $\tau_{j,s} \leq \tau_{j,e} = \tau_{j+1,s}$. Each offer has a unique index i. The distinction between IL or SR is not made by i, rather create a function $OFF(i) : \mathbb{N} \to \{IL, SR\}$ to contain that information.

For each region and offer there is a maximum amount of reserve that can be offered, $\Delta_{i,j}$. For SR, $\Delta_{i,j} = g_i(\tau_{j,e} - \tau_{j,s})$ if the SR offer can ramp up in that time period. For

IL, $\Delta_{i,j} = p_i^{max}$ in the region it is initiated. For some regions there are no SR or IL offers, call these ones NN. A function is created to contain the information on the type of region: $REG(j) : \mathbb{N} \to \{SR, IL, NN\}$. From the definitions so far, four rules are implied:

1. It is not possible to have two consecutive IL regions, i.e. $REG(j) = IL$ and $REG(j+1) = IL$, because that would imply more than one IL offer with the same time, $t_{i,p}$, which is not possible, because all multiple instances have been combined into one.

2. It is not possible to have two consecutive NN regions, i.e. $REG(j) = NN$ and $REG(j+1) = NN$, as this would imply that $\tau_{j,e} \in \Omega$ which it is not.

3. If there are consecutive SR regions then there exists a SR offer that has either finished ramping by the end of the first region, or a SR offer that has started ramping at the start of the second region.

4. If $OFF(i) = IL$, $REG(j) = IL$ and $\Delta_{i,j} > 0$, then $\forall k \neq i, \Delta_{k,j} = 0$ and $\forall k \neq j, \Delta_{i,k} = 0$. This is the exclusivity of an IL offer to one region, and no other offer in that same region.

The next step is to introduce a series of variables, $\mu_{i,j}(\lambda) \in [0, 1]$, which define the dispatch of each offer in each region:

$$u_k = \sum_{j=1}^{Nreg} \mu_{i,j} \Delta_{i,j} \tag{A.2}$$

where i and k refer to the same SR offer and $Nreg$ are the number of regions. A similar equation exists for p_k. To remove any ambiguity, if $\Delta_{i,j} = 0$ then $\mu_{i,j} = 0$ for all λ. A restriction is placed on $\mu_{i,j}$ when $OFF(i) = SR$:

1. If $0 < \mu_{i,j} < 1$ then $\forall k < j$ and $\Delta_{i,k} > 0$ it is required that $\mu_{i,k} = 1$, and $\forall k > j$ it is required that $\mu_{i,k} = 0$.

2. If $\mu_{i,j} = 1$ and the next possible region $l > j$, where $\Delta_{i,l} > 0$, $\mu_{i,l} = 0$, then $\forall k < j$ and $\Delta_{i,k} > 0$ it is required that $\mu_{i,k} = 1$, also $\forall k > l$ it is required that $\mu_{i,k} = 0$.

This is to recognise that each SR offer ramps in a continuous sequence, not stopping and starting. The next step is to understand how $\mu_{i,j}$ changes with λ as $\mathbf{v}_1$ transitions to $\mathbf{v}_2$. For IL offers, assuming $REG(j) = IL$ and $\Delta_{i,j} > 0$:

$$\mu_{i,j}(\lambda) = \frac{v_{i,1}}{\Delta_{i,j}} + \lambda \frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} \tag{A.3}$$

For each SR offer the critical region has to be found, where $0 < \mu_{i,j} < 1$. Define $\theta_{i,j,s}$ and $\theta_{i,j,e}$, the amount of reserve dispatched at the start and end of each region:

$$\theta_{i,j,s} = \sum_{k=1}^{j-1} \Delta_{i,k} \quad \text{and} \quad \theta_{i,j,e} = \theta_{i,j,s} + \Delta_{i,j} \tag{A.4}$$

The critical region is the one where $\theta_{i,j,s} < (1-\lambda)v_{i,1} + \lambda v_{i,2} < \theta_{i,j,e}$:

$$\mu_{i,j}(\lambda) = \frac{v_{i,1} - \theta_{i,j,s}}{\Delta_{i,j}} + \lambda \frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} \tag{A.5}$$

A.3 TIME OF THE MINIMUM FREQUENCY

The time the minimum frequency occurs is when total reserve equals the risk, and more precisely described in the conditions of Eqs. 5.9 and 5.10. Therefore create a series of variables $\gamma_j(\lambda) \in [0,1]$ that determines when that is:

1. Boundary Situation (BS), if $\gamma_{j-1} = \max\{\mu_{i,j-1}\}_i$ and $\gamma_j = 0$ then for all $k < j-1$, $\gamma_k = 1$ and for all $k > j$, $\gamma_k = 0$, the total reserve is:

$$S(\gamma) = \sum_{k=1}^{j-1} \sum_{i=1}^{Noff} \mu_{i,k} \Delta_{i,k} \tag{A.6}$$

2. Gap Situation (GS), if $0 < \gamma_j < \max\{\mu_{i,j}\}_i$ then for all $k < j$, $\gamma_k = 1$ and for all $k > j$, $\gamma_k = 0$, the total reserve as a function of γ is:

$$S(\gamma) = \sum_{k=1}^{j-1} \sum_{i=1}^{Noff} \mu_{i,k} \Delta_{i,k} + \sum_{\substack{i=1 \\ \mu_{i,j} \leq \gamma_j}}^{Noff} \mu_{i,j} \Delta_{i,j} + \sum_{\substack{i=1 \\ \mu_{i,j} > \gamma_j}}^{Noff} \gamma_j \Delta_{i,j} \tag{A.7}$$

where γ is called the position marker:

$$\gamma = \sum_{j=1}^{Nreg} \gamma_j \tag{A.8}$$

For γ to determine the minimum time:

$$(1-\lambda)R_1 + \lambda R_2 = S(\gamma) \tag{A.9}$$

and for γ to be unique, another condition is placed on δ such that for all $0 \leq \delta < \gamma$:

$$(1-\lambda)R_1 + \lambda R_2 > S(\delta) \tag{A.10}$$

The minimum time is:

$$t_{min} = \begin{cases} (1-\gamma_{j-1})\tau_{j-1,s} + \gamma_{j-1}\tau_{j-1,e} & \text{BS} \\ (1-\gamma_j)\tau_{j,s} + \gamma_j\tau_{j,e} & \text{GS} \end{cases} \tag{A.11}$$

where j is the same value applied in Eqs. A.6 and A.7.

The dispatched reserve up to the minimum time as defined in Eq. 5.44 and 5.45, q_l and w_l, are:

$$q_l = \begin{cases} \sum_{k=1}^{j-1} \mu_{i,k}\Delta_{i,k} & \text{BS} \\ \left(\sum_{k=1}^{j-1} \mu_{i,k}\Delta_{i,k}\right) + \mu_{i,j}\Delta_{i,j} & \text{GS and } \mu_{i,j} \leq \gamma_j \\ \left(\sum_{k=1}^{j-1} \mu_{i,k}\Delta_{i,k}\right) + \gamma_j\Delta_{i,j} & \text{GS and } \mu_{i,j} > \gamma_j \end{cases} \tag{A.12}$$

where the indices i and l refer to the same offer, but not necessarily to the same number. This is because i in this Appendix can uniquely define whether it is an IL or SR offer by its value, but the value of l requires further information to distinguish it from IL or SR. This information is found in whether the symbol 'q' or 'w' is being used. The equation for w_l is the same as for q_l in Eq. A.12, except q_l is replaced with w_l.

A.4 PROGRESSION OF MINIMUM TIME

This section describes the progression of minimum time with changes in λ. The domain $\lambda \in [0,1]$ is divided into states, $(\lambda_{T-1}, \lambda_T)$ is partitioned to each, with transitions between states at λ_T. There are three different types of states to consider:

1. Boundary State, $BDY(S,E)$: $\gamma_S = \max\{\mu_{i,S}\}_i > 0$ and $\gamma_E = 0$, where either $S+1 = E$, or $S+2 = E$ with $REG(S+1) = NN$, or $\forall i, l$ where $S < l < E$ such that $\mu_{i,l} = 0$. S and E are the regions the boundary is in between. For this to be a state:

$$R_1 = \sum_{i \in D} v_{i,1} + \sum_{i \in E} \theta_{i,S,e} \tag{A.13}$$

$$R_2 - R_1 = \sum_{i \in D} v_{i,2} - v_{i,1} \tag{A.14}$$

where

$$D = \{i \quad \text{where} \quad \exists j \leq S \quad \text{s.t.} \quad \mu_{i,j} > 0, \quad \text{if} \quad \mu_{i,S} = 1 \quad \text{then} \quad v_{i,2} - v_{i,1} < 0\}$$

$$E = \{i \quad \text{where} \quad \mu_{i,S} = 1 \quad \text{and} \quad (\exists j \geq E \quad \text{s.t.} \quad \mu_{i,j} > 0 \quad \text{or} \quad v_{i,2} - v_{i,1} \geq 0)\}$$

for all λ in the domain $(\lambda_{T-1}, \lambda_T)$. The sets D and E are the offers that stop ramping before the minimum time and the SR offers that continue ramping past

the minimum time.

2. IL Gap State, $ILG(j)$: $REG(j) = IL$, $\exists i$ where $\Delta_{i,j} > 0$, and $0 < \gamma_j < \mu_{i,j}$.

3. SR Gap State, $SRG(j)$: $REG(j) = SR$ and $0 < \gamma_j < \max\{\mu_{i,j}\}_i$.

The next stage is to describe all possible changes that could occur to each state, including transitions to other states. The possible changes are organised by the initial state, and then by how it progresses with λ. Each possible change is named in the following convention: *Initial State* to *Transition Type* to *Final State*, thereby keeping record of each situation.

A.4.1 $SRG(j)$ Transitions

Initially when the last amount of risk is covered by SR, the change in position marker, $\gamma_j(\lambda)$, is derived by substituting Eq. A.7 into Eq. A.9 and rearranging for γ_j, but first the substitution:

$$(1-\lambda)R_1 + \lambda R_2 = \sum_{k=1}^{j-1}\sum_{i=1}^{Noff} \mu_{i,k}\Delta_{i,k} + \sum_{\substack{i=1 \\ \mu_{i,j}\leq\gamma_j}}^{Noff} \mu_{i,j}\Delta_{i,j} + \sum_{\substack{i=1 \\ \mu_{i,j}>\gamma_j}}^{Noff} \gamma_j\Delta_{i,j} \tag{A.15}$$

The choice of region j is determined by finding $0 < \gamma_j < 1$. Substituting Eq. A.3 and A.5 into Eq. A.15:

$$(1-\lambda)R_1 + \lambda R_2 = \sum_{i\in D}(v_{i,1} + \lambda(v_{i,2} - v_{i,1})) + \sum_{i\in E}(\theta_{i,j,s} + \gamma_j\Delta_{i,j}) \tag{A.16}$$

where $\quad D = \{i \quad \text{where} \quad \mu_{i,j} \leq \gamma_j, \quad \text{if} \quad \mu_{i,j} = 0 \quad \exists l < j \quad \text{s.t.} \quad \mu_{i,l} > 0\}$

$$E = \{i \quad \text{where} \quad \mu_{i,j} > \gamma_j\}$$

rearranging for γ_j:

$$\gamma_j = \frac{B_{l,0} + B_{l,1}\lambda}{C_l} \tag{A.17}$$

where

$$B_{l,0} = R_1 - \sum_{i\in D} v_{i,1} - \sum_{i\in E}\theta_{i,j,s}$$

$$B_{l,1} = R_2 - R_1 - \sum_{i\in D}(v_{i,2} - v_{i,1}) \quad \text{and} \quad C_l = \sum_{i\in E}\Delta_{i,j}$$

The index l marks the changes in $B_{l,0}$, $B_{l,1}$, and C_l as transitions occur, and is separate to its use in Eq. A.16. The progress of γ_j is dependent on the sign of $B_{l,1}$, if positive then the minimum time increases. There are two possible scenarios that do not change the state:

1. One SR offer which initially stopped ramping before t_{min}, has stopped ramping on t_{min}, i.e. $\exists i$ s.t. for $\lambda < \lambda_T$, $\mu_{i,j} < \gamma_j$ but when $\lambda = \lambda_T$, $\mu_{i,j} = \gamma_j$. The following condition is required for $\mu_{i,j}$ to intersect γ_j:

$$\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} > \frac{B_{l,1}}{C_l} \tag{A.18}$$

2. The opposite of scenario 1, one SR offer that had stopped ramping after t_{min} now stops on it at λ_T. The opposite condition is required:

$$\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} < \frac{B_{l,1}}{C_l} \tag{A.19}$$

For a single instance of scenario 1, initially $\mu_{i,j} < \gamma_j$, they became equal, and they have to swap $\mu_{i,j} > \gamma_j$. Once swapped the function for γ_j changes:

$$\gamma_j = \frac{B_{l+1,0} + B_{l+1,1}\lambda}{C_{l+1}} \tag{A.20}$$

where

$$B_{l+1,0} = B_{l,0} + v_{i,1} - \theta_{i,j,s} \quad \text{and} \quad B_{l+1,1} = B_{l,1} + v_{i,2} - v_{i,1} \quad \text{and} \quad C_{l+1} = C_l + \Delta_{i,j}$$

It needs to be shown that $\mu_{i,j}$ still surpasses γ_j after the change, i.e.:

$$\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} > \frac{B_{l+1,1}}{C_{l+1}} \tag{A.21}$$

starting with Eq. A.18, multiply both sides by both denominators:

$$C_l(v_{i,2} - v_{i,1}) > \Delta_{i,j} B_{l,1} \tag{A.22}$$

add $\Delta_{i,j}(v_{i,2} - v_{i,1})$ to both sides:

$$(C_l + \Delta_{i,j})(v_{i,2} - v_{i,1}) > \Delta_{i,j}(B_{l,1} + v_{i,2} - v_{i,1})$$

and dividing both sides by C_{l+1} and $\Delta_{i,j}$ it arrives at Eq. A.21. It is not possible for $C_l = 0$ as this would imply a boundary had initially been reached. It will become

important later to understand how $B_{l,1}/C_l$ compares with $B_{l+1,1}/C_{l+1}$. Starting with the relationships between l and $l+1$:

$$\frac{B_{l,1}}{C_l} = \frac{B_{l+1,1} - (v_{i,2} - v_{i,1})}{C_{l+1} - \Delta_{i,j}} \tag{A.23}$$

multiplying by both denominators, and shifting one of the terms from the left to the right side:

$$B_{l,1}C_{l+1} = B_{l+1,1}C_l + B_{l,1}\Delta_{i,j} - C_l(v_{i,2} - v_{i,1}) \tag{A.24}$$

from Eq. A.22, $\Delta_{i,j}B_{l,1} - C_l(v_{i,2} - v_{i,1}) < 0$, therefore $B_{l,1}C_{l+1} < B_{l+1,1}C_l$ and:

$$\frac{B_{l+1,1}}{C_{l+1}} > \frac{B_{l,1}}{C_l} \tag{A.25}$$

The rate at which γ_j changes with λ increases across the transition. Now considering a single instance of scenario 2, SR that stopped ramping after the minimum frequency now stops before, and the change in γ_j changes as it does in Eq. A.20, but the transformation in variables is reversed:

$$B_{l+1,0} = B_{l,0} - w_{i,1} + \theta_{i,j,s} \quad \text{and} \quad B_{l+1,1} = B_{l,1} - (w_{i,2} - w_{i,1}) \quad \text{and} \quad C_{l+1} = C_l - \Delta_{i,j}$$

After the transition it is true that $\mu_{i,j} < \gamma_j$, as:

$$\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} < \frac{B_{l+1,1}}{C_{l+1}} \tag{A.26}$$

the working is not shown but follows the same process as for scenario 1. More interesting is the relationship between $B_{l,1}/C_l$ and $B_{l+1,1}/C_{l+1}$, and starting with the relationship between l and $l+1$:

$$\frac{B_{l,1}}{C_l} = \frac{B_{l+1,1} + (v_{i,2} - v_{i,1})}{C_{l+1} + \Delta_{i,j}} \tag{A.27}$$

$$B_{l,1}C_{l+1} = B_{l+1,1}C_l + C_l(v_{i,2} - v_{i,1}) - B_{l,1}\Delta_{i,j} \tag{A.28}$$

but this time from Eq. A.19, $C_l(v_{i,2} - v_{i,1}) - B_{l,1}\Delta_{i,j} < 0$:

$$\frac{B_{l+1,1}}{C_{l+1}} > \frac{B_{l,1}}{C_l} \tag{A.29}$$

the same as Eq. A.25, this seems to imply Eq. A.25 is always true independent of the number of crossings that occur at λ_T, i.e. multiple coincidences of scenario 1 and 2. To

show this, define the set D_{BA} where each SR offer follows scenario 1, and E_{AB} which follows scenario 2. The general transformation formulas are:

$$B_{l+1,0} = B_{l,0} + \sum_{i \in D_{BA}} (v_{i,1} - \theta_{i,j,s}) - \sum_{i \in E_{AB}} (v_{i,1} - \theta_{i,j,s}) \tag{A.30}$$

$$B_{l+1,1} = B_{l,1} + \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) - \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}) \tag{A.31}$$

$$C_{l+1} = C_l + \sum_{i \in D_{BA}} \Delta_{i,j} - \sum_{i \in E_{AB}} \Delta_{i,j} \tag{A.32}$$

Following the working presented previously for Eq. A.27:

$$\frac{B_{l,1}}{C_l} = \frac{B_{l+1,1} - \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1})}{C_{l+1} - \sum_{i \in D_{BA}} \Delta_{i,j} + \sum_{i \in E_{AB}} \Delta_{i,j}} \tag{A.33}$$

$$B_{l,1}C_{l+1} = B_{l+1,1}C_l + B_{l,1}(\sum_{i \in D_{BA}} \Delta_{i,j} - \sum_{i \in E_{AB}} \Delta_{i,j}) + C_l(-\sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}))$$

$$B_{l,1}C_{l+1} = B_{l+1,1}C_l + \sum_{i \in D_{BA}} (B_{l,1}\Delta_{i,j} - C_l(v_{i,2} - v_{i,1})) + \sum_{i \in E_{AB}} (C_l(v_{i,2} - v_{i,1}) - B_{l,1}\Delta_{i,j})$$

from the conditions of scenario 1 and 2:

$$\frac{B_{l+1,1}}{C_{l+1}} > \frac{B_{l,1}}{C_l}$$

It has to be shown whether Eq. A.21 and A.26 still hold. For scenario 2, Eq. A.26 always holds. Once SR offers transition from stop ramping after the minimum time to before, it cannot transition back again before γ_j reaches the boundary. The same is not necessarily true for scenario 1, as it is now possible for:

$$\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} < \frac{B_{l+1,1}}{C_{l+1}} \tag{A.34}$$

i.e. some SR do not transition fully, as $\mu_{i,j} < \gamma_j$ remains. It will not be shown which ones do and do not fully transition, but a condition has to be satisfied. Define the set $D_S \subset D_{BA}$ which does not transition and $D_T \subseteq D_{BA}$ that does. Adjusting the transformations of Eq. A.30, A.31, and A.32 so that D_{BA} is replaced by D_T, the condition is:

$$\forall i \in D_S \quad \frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} \leq \frac{B_{l+1,1}}{C_{l+1}} \tag{A.35}$$

For these general transitions give the name $T\_SR(j)$. Therefore by the naming convention of transitions types, the longer description is $SRG(j)$ to $T\_SR(j)$ to $SRG(j)$. For transitions that change the state, there are three possible types:

1. Reach the bottom of the region, $\gamma_j(\lambda_T) = 0$.
2. Reach the top of the region, $\gamma_j(\lambda_T) = \max\{\mu_{i,j}\}_i$
3. Both 1 and 2.

Consider type 1. Firstly, γ_j is decreasing to reach zero, $B_{l,1}/C_l < 0$. When $\gamma_j = 0$ what happens next is dependent on the amount of reserve dispatched in the regions up to $j-1$ and whether this is enough to cover the change in risk. The total amount of reserve in those regions for $\lambda > \lambda_T$ is:

$$\sum_{i \in D} (v_{i,1} + \lambda(v_{i,2} - v_{i,1})) + \sum_{i \in E} \theta_{i,j,s} \tag{A.36}$$

$$D = \{i \quad \text{where} \quad \exists l \leq j-1 \quad \text{s.t.} \quad \mu_{i,l}(\lambda_T) > 0, \quad \text{if} \quad \mu_{i,j-1}(\lambda_T) = 1$$

$$\text{then} \quad v_{i,2} - v_{i,1} < 0 \quad \text{and} \quad \forall k > j-1, \quad \mu_{i,k}(\lambda_T) = 0\}$$

$$E = \{i \quad \text{where} \quad \mu_{i,j-1}(\lambda_T) = 1 \quad \text{and} \quad (\exists k > j-1 \quad \text{s.t.} \quad \mu_{i,k}(\lambda_T) > 0 \quad \text{or} \quad v_{i,2} - v_{i,1} \geq 0)\}$$

when $\gamma_j(\lambda_T) = 0$. The first summation is from reserve that stops ramping before the minimum time. There are three possibilities:

1. $R_2 - R_1 > \sum_{i \in D} v_{i,2} - v_{i,1}$, risk is increasing at a greater rate than reserve and so have to go back to $SRG(j)$ state. The transition is called $LOW\_SR(j)$. The progression is:

   $SRG(j)$ to $LOW\_SR(j)$ to $SRG(j)$

2. $R_2 - R_1 = \sum_{i \in D} v_{i,2} - v_{i,1}$, the position marker remains on the boundary, the transition is called $HIGH\_BYD(j)$.

   $SRG(j)$ to $HIGH\_BDY(j)$ to $BDY(j-1, j)$

3. $R_2 - R_1 < \sum_{i \in D} v_{i,2} - v_{i,1}$, more reserve than risk, minimum time reduces and position marker drops to the region below. The general transition between regions is called $T\_BDY(j-1, j)$. It can either drop to SR or IL in the region below.

   $SRG(j)$ to $T\_BDY(j-1, j)$ to $SRG(j-1)$

   $SRG(j)$ to $T\_BDY(j-1, j)$ to $ILG(j-1)$

For type 2 and 3 events there is only one possibility, the state changes to the next region above (meaning later in time):

1. $SRG(j)$ to $T\_BDY(j, j+1)$ to $SRG(j+1)$ or $ILG(j+1)$ for type 2
2. $SRG(j)$ to $T\_BDY(j-1, j+1)$ to $SRG(j+1)$ or $ILG(j+1)$ for type 3

It has to be shown that this is the only possibility. This is done separately for $\max\{\mu_{i,j}\}_i < 1$ and then when $\max\{\mu_{i,j}\}_i = 1$ when $\lambda = \lambda_T$. For the first case there are SR offers, i, where $\mu_{i,j} > \gamma_j$ before $\lambda = \lambda_T$, and when equal then $\mu_{i,j}(\lambda_T) = \gamma_j(\lambda_T) = \max\{\mu_{i,j}(\lambda_T)\}_i$. These offers are combined in the set D_{AB}. It is known that:

$$\forall i \in D_{AB} \quad \frac{v_{i,2} - v_{i,1}}{\Delta_{i,j}} < \frac{B_{l,1}}{C_l} = \frac{(R_2 - R_1) - \sum_{k \in D_{BB}} (v_{k,2} - v_{k,1})}{\sum_{k \in D_{AB}} \Delta_{k,j}} \tag{A.37}$$

where D_{BB} are all the initial offers that finish ramping before t_{min}. For the above transition to be the only possibility it has to be proven that the amount of reserve in all the blocks up to j is strictly less than the risk, for $\lambda > \lambda_T$, or equivalently:

$$R_2 - R_1 > \sum_{i \in D} (v_{i,2} - v_{i,1}) \tag{A.38}$$

where

$$D = \{i \quad \text{where} \quad \exists l \leq j \quad \text{s.t.} \quad \mu_{i,l}(\lambda_T) > 0, \quad \text{if} \quad \mu_{i,j}(\lambda_T) = 1$$

$$\text{then} \quad v_{i,2} - v_{i,1} < 0 \quad \text{and} \quad \forall k > j, \quad \mu_{i,k}(\lambda_T) = 0\}$$

$D = D_{BB} \cup D_{AB}$, D_{AB} is combined with D_{BB} as the reserve that stops ramping before minimum time, and the extra condition when $\mu_{i,j}(\lambda_T) = 1$ is not considered as $\max\{\mu_{i,j}(\lambda_T)\}_i < 1$. To arrive at Eq. A.38, Eq. A.37 is rearranged by multiplying by both denominators:

$$(v_{i,2} - v_{i,1}) \sum_{k \in D_{AB}} \Delta_{k,j} < \Delta_{i,j} \Big((R_2 - R_1) - \sum_{k \in D_{BB}} (v_{k,2} - v_{k,1}) \Big) \tag{A.39}$$

Since there is an individual equation for each $i \in D_{AB}$, adding them all together:

$$\Big(\sum_{i \in D_{AB}} v_{i,2} - v_{i,1} \Big) \Big(\sum_{k \in D_{AB}} \Delta_{k,j} \Big) < \Big(\sum_{i \in D_{AB}} \Delta_{i,j} \Big) \Big((R_2 - R_1) - \sum_{k \in D_{BB}} v_{k,2} - v_{k,1} \Big) \tag{A.40}$$

Dividing by the common term on both sides, and shifting the summation of $v_{k,2} - v_{k,1}$ to the left hand side, Eq. A.38 is achieved. Now for the second case when $\max\{\mu_{i,j}(\lambda_T)\}_i = 1$, again Eq. A.38 has to be proven. It is known that $B_{l,1}/C_l > 0$ in order for $\gamma_j = 1$ at λ_T, consequently:

$$R_2 - R_1 > \sum_{i \in D_0} v_{i,2} - v_{i,1} \tag{A.41}$$

where

$$D_0 = \{i \quad \text{where} \quad \mu_{i,j} \leq \gamma_j, \quad \text{if} \quad \mu_{i,j} = 0 \quad \exists l < j \quad \text{s.t.} \quad \mu_{i,l} > 0\}$$

To arrive at D of Eq. A.38, the SR offers that catch up to γ_j need to be removed, i.e the set D_{BA}; there is also a set of SR offers, E_{AB}, that need to be added because they satisfy the $\mu_{i,j}(\lambda_T) = 1$ condition of D. These are SR offers where for $\lambda > \lambda_T$, $\mu_{i,j}(\lambda) < 1$, but initial $\mu_{i,j} = 1$ for $\lambda \leq \lambda_T$. Therefore $D = (D_0 \setminus D_{BA}) \cup E_{AB}$. From the above description it is known that:

$$\forall i \in D_{BA} \quad v_{i,2} - v_{i,1} > 0 \quad \text{and} \quad \forall i \in E_{AB} \quad v_{i,2} - v_{i,1} < 0 \tag{A.42}$$

Therefore subtracting each D_{BA} term from Eq. A.41 on the right hand side, and adding each E_{AB} term also, Eq. A.38 is proven correct. It is important to note for these transitions that travel up a region, which will be helpful in later analysis, is that γ_{j+1} is increasing with λ in the next state, as Eq. A.38 shows.

A.4.2 $ILG(j)$ Transitions

When an IL offer is the last offer to cover the risk, the changes in position marker are found by substituting Eq. A.7 into Eq. A.9 as was done for $SRG(j)$ transitions:

$$(1-\lambda)R_1 + \lambda R_2 = \gamma_j \Delta_{p,j} + \sum_{k=1}^{j-1} \sum_{i=1}^{Noff} \mu_{i,k} \Delta_{i,k} \tag{A.43}$$

where p is the active IL offer for region j. Substituting in Eq. A.3 and A.5 into Eq. A.43:

$$(1-\lambda)R_1 + \lambda R_2 = \gamma_j \Delta_{p,j} + \sum_{i \in D} (v_{i,1} + \lambda(v_{i,2} - v_{i,1})) + \sum_{i \in E} \theta_{i,j-1,e} \tag{A.44}$$

where

$$D = \{i \quad \text{where} \quad \mu_{i,j-1} \leq 1, \quad \text{if} \quad \mu_{i,j-1} = 0 \quad \exists l < j-1 \quad \text{s.t.} \quad \mu_{i,l} > 0,$$

$$\text{if} \quad \mu_{i,j-1} = 1 \quad \text{then} \quad v_{i,2} - v_{i,1} < 0 \quad \text{and} \quad \mu_{i,j+1} = 0\}$$

$$E = \{i \quad \text{where} \quad \mu_{i,j-1} = 1 \quad \text{and either} \quad v_{i,2} - v_{i,1} \geq 0 \quad \text{or} \quad \mu_{i,j+1} > 0\}$$

rearranging for γ_j:

$$\gamma_j = \frac{B_{l,0} + B_{l,1}\lambda}{\Delta_{p,j}} \tag{A.45}$$

where

$$B_{l,0} = R_1 - \sum_{i \in D} v_{i,1} - \sum_{i \in E} \theta_{i,j-1,e}$$

$$B_{l,1} = R_2 - R_1 - \sum_{i \in D} (v_{i,2} - v_{i,1})$$

Similar to the $SRG(j)$ state there are two possible scenarios that do not change the state, but only if $REG(j-1) = SR$, $REG(j+1) = SR$, and $\tau_{j-1,e} = \tau_{j,s} = \tau_{j,e} = \tau_{j+1,s}$:

1. SR initially stops ramping before the time of the IL offer, p, but then stops after.
2. The opposite of above, SR initially stopping after the IL offer, p, but transitions to stopping before.

The important thing to understand is how the function for γ_j changes, consider the first case, there has to be an offer i where $\mu_{i,j-1}(\lambda_T) = 1$ and for $\lambda < \lambda_T$, $\mu_{i,j-1}(\lambda) < 1$. Therefore $v_{i,2} - v_{i,1} > 0$, and the changes in the function are:

$$B_{l+1,0} = B_{l,0} + v_{i,1} - \theta_{i,j-1,e} \tag{A.46}$$

$$B_{l+1,1} = B_{l,1} + (v_{i,2} - v_{i,1}) \tag{A.47}$$

and so $B_{l+1,1} > B_{l,1}$. For the second case when $\lambda > \lambda_T$, $\mu_{i,j-1}(\lambda) < 1$, then $v_{i,2} - v_{i,1} < 0$. The changes in the function for γ_j are:

$$B_{l+1,0} = B_{l,0} - v_{i,1} + \theta_{i,j-1,e} \tag{A.48}$$

$$B_{l+1,1} = B_{l,1} - (v_{i,2} - v_{i,1}) \tag{A.49}$$

and again $B_{l+1,1} > B_{l,1}$. Therefore each time these transitions occur, the rate at which γ_j changes with λ is always increasing. This transition is called $ILG(j)$ to $T\_IL(j)$ to $ILG(j)$. For transitions that change the state there are:

1. Reaching the bottom of the region, $\gamma_j(\lambda_T) = 0$.
2. Reaching the top of the region, $\gamma_j(\lambda_T) = \mu_{p,j}$

3. Both can happen together like it did in $SRG(j)$ based transitions, but this can only occur when $\lambda = 1$, and will not be considered further.

Following a similar reasoning for $SRG(j)$, for type 1 events there are three possibilities:

1. $ILG(j)$ to $LOW\_IL(j)$ to $ILG(j)$. This one and the next are only possible when there are SR offers that initially stop ramping before $\tau_{j,s}$, but at λ_T stop ramping at the time of the critical IL offer, $\tau_{j,s}$.
2. $ILG(j)$ to $HIGH\_BDY(j-1,j)$ to $BDY(j-1,j)$
3. $ILG(j)$ to $T\_BDY(j-1,j)$ to $SRG(j-1)$ (or $ILG(j-1)$)

For type 2 the only transition is to go to the next region, $ILG(j)$ to $T\_BDY(j,j+1)$ to $SRG(j+1)$ (or $ILG(j+2)$). In the next state γ_{j+1} is increasing with λ.

A.4.3 $BDY(j-1,j)$ Transitions

The situation where the reserve in all the regions up to and including $j-1$ satisfies the risk requirement for a range of λ values. The amount of reserve in the first $j-1$ regions is:

$$\sum_{k=1}^{j-1}\sum_{i=1}^{Noff}\mu_{i,k}\Delta_{i,k} = \sum_{i\in D}(v_{i,1}+\lambda(v_{i,2}-v_{i,1})) + \sum_{i\in E}\theta_{i,j-1,e} \tag{A.50}$$

where D and E are the same as in Eq. A.44. For the total reserve to equal the risk over a range of λ it is necessary that:

$$R_1 = \sum_{i\in D} v_{i,1} + \sum_{i\in E}\theta_{i,j-1,e} \tag{A.51}$$

$$R_2 - R_1 = \sum_{i\in D} v_{i,2} - v_{i,1} \tag{A.52}$$

as stated in Eq. A.13 and A.14. There are two possible scenarios that could change the state: a SR offer which initially stopped ramping before the boundary of $j-1$ and j, now stops ramping after the boundary, and vice versa. The summation term of Eq. A.52 is captured by the following parameter for ease of reference:

$$B_{l,2} = \sum_{i\in D} v_{i,2} - v_{i,1} \tag{A.53}$$

For the first case $v_{i,2} - v_{i,1} > 0$ so that $\mu_{i,j-1}(\lambda_T) = 1$, when initially it was less than one. Therefore $B_{l+1,2} = B_{l,2} - (v_{i,2} - v_{i,1}) < R_2 - R_1$, and the state goes to the next

region up. For the second case $v_{i,2} - v_{i,1} < 0$, then $B_{l+1,2} = B_{l,2} + (v_{i,2} - v_{i,1}) < R_2 - R_1$, and transitions into the next region above. Therefore there is only one transition direction possible out of $BDY(j-1, j)$:

$BDY(j-1, j)$ to $HIGH\_BDY(j-1, j)$ to $SRG(j)$ (or $ILG(j)$)

For the $SRG(j)$ and $ILG(j)$ states the minimum time is connected with the change in γ_j. However for the $BDY(j-1, j)$ state the minimum time is connected with $\max\{\mu_{i,j-1}\}_i$ as well, this complexity needs further explanation. In most incidents of the $BDY(j-1, j)$ state $\max\{\mu_{i,j-1}\}_i = 1$ or $REG(j-1) = IL$, so $t_{min} = \tau_{j,s}$ for the whole period, but there are situations where this does not happen and t_{min} varies. The rest of this section describes this variation.

In the $BDY(j-1, j)$ state, SR offer k has the maximum dispatch, i.e. $\mu_{k,j-1} = \max\{\mu_{i,j-1}\}_i$. If there are multiple incidents of dispatches equalling the maximum, choose the offer with the largest $(v_{i,2} - v_{i,1})/\Delta_{i,j-1}$. To simplify the notation express $B_{k,3} = (v_{k,2} - v_{k,1})/\Delta_{k,j-1}$. If $B_{k,3}$ is positive then t_{min} is increasing. The maximum offer changes by:

1. another SR offer, $k+1$, replacing the original as being the maximum dispatch. For this to happen $B_{k+1,3} > B_{k,3}$, as initially $\mu_{k+1,j-1} < \mu_{k,j-1}$.

2. SR offer, k, reaching the bottom of the region, $\mu_{k,j-1} = 0$, and the region below, $REG(j-2) = IL$. When this happens require $B_{k+1,3} = 0$ as $t_{min} = \tau_{j-1,s}$ and is constant. The minimum time is unchanged until another SR offer, $k+2$, increases past $\tau_{j-1,s}$ which requires $B_{k+2,3} > 0$.

3. reaching the bottom of the region and $REG(j-2) = SR$. This time the maximum offer does not change, but is a possibility.

These transitions are called $TMIN\_CHG$:

$BDY(j-1, j)$ to $TMIN\_CHG$ to $BDY(j-1, j)$

A.5 THE FIRST MAIN RESULT - TIME OF MINIMUM FREQUENCY

Throughout analysing all the transitions, a lack of symmetry is noticed. It is not possible to go down one region once already having gone up one, but in the reverse it is possible. This result is generalised for its impact on the time the minimum frequency occurs, $t_{min}(\lambda)$.

Result 1: *For all* $\lambda_1, \lambda_2 \in [0, 1)$ *where* $0 \leq \lambda_1 < \lambda_2 < 1$, *if* $t_{min}(\lambda_1) < t_{min}(\lambda_2)$ *then for all* $\alpha, \beta \in [\lambda_2, 1]$ *where* $\lambda_2 \leq \alpha \leq \beta \leq 1$, $t_{min}(\alpha) \leq t_{min}(\beta)$. *That is when* t_{min} *starts to increase it will always be monotonically increasing after that.*

The first part of the proof is to show for $\lambda_1 < \lambda_2$ in the same state, where $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$, without any transitions in between, that the minimum time will always be non-decreasing for $\lambda \geq \lambda_2$. There is one exception when in the $ILG(j)$ state, one extra condition is required.

1. $SRG(j)$ State

   For $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$ it requires $B_{l,1}/C_l \geq 0$. There are three possibilities for progression in λ:

   - $\lambda = 1$ before any transitions occur. This is the trivial case because $B_{l,1}/C_l \geq 0$, and $t_{min}(\lambda)$ will always be non-decreasing during this time.
   - A $T\_SR(j)$ transition occurs. It is still non-decreasing past the transition because $B_{l+1,1}/C_{l+1} > B_{l,1}/C_l \geq 0$. After this consider the three possibilities again until it terminates.
   - A $T\_BDY(j, j+1)$ transition occurs. $B_{l+1,1}/C_{l+1}$ is positive in $SRG(j+1)$ or $ILG(j+1)$ because:

   $$R_2 - R_1 - \sum_{i \in D} v_{i,2} - v_{i,1} > 0 \tag{A.54}$$

   where

   $$D = \{i \quad \text{where} \quad \mu_{i,j} \leq 1, \quad \text{if} \quad \mu_{i,j} = 0 \quad \exists l < j \quad \text{s.t.} \quad \mu_{i,l} > 0,$$

   $$\text{if} \quad \mu_{i,j} = 1 \quad \text{then} \quad v_{i,2} - v_{i,1} < 0 \quad \text{and} \quad \mu_{i,l} = 0 \quad \forall l > j\}$$

   That is the amount of reserve in the regions up to j is decreasing, and has to be compensated by increasing γ_{j+1} and consequently the minimum time. From here reconsider the three possibilities. If $ILG(j+1)$ then the ones for the $ILG(j)$ state.

2. $ILG(j)$ State

   For this state an extra condition is required, because for any two points in this state $t_{min}(\lambda_1) = t_{min}(\lambda_2) = \tau_{j,s}$, and it is still possible for the minimum time to decrease through $\gamma_j = 0$. The extra condition is $B_{l,1} \geq 0$. There are now three possibilities:

- $\lambda = 1$ before any transitions occur. Trivial because $t_{min}(\lambda) = \tau_{j,s}$, the minimum time does not change.
- A $T\_IL(j)$ transition occurs. Minimum time remains constant and $B_{l+1,1} > B_{l,1} \geq 0$. Check the three possibilities again for the next option.
- A $T\_BDY(j, j+1)$ transition occurs. For the same reason for the $SRG(j)$ state, $B_{l+1,1}/C_{l+1} > 0$. Check again the possibilities until it terminates.

3. $BDY(j, j+1)$ State

   This state is different than the previous two. It can only be in this state once over all j. It always exits to $SRG(j+1)$ or $ILG(j+1)$, γ_{j+1} is always increasing, so continue with the possibilities of the IL and SR states once this happens. Unless it terminates at $\lambda = 1$ first. However there could be several internal transitions, $TMIN\_CHG$. For $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$, $B_{k,3} > 0$, and since $B_{k+1,3} > B_{k,3}$ the minimum time is always increasing.

The second part is to show for $\lambda_1 < \lambda_2$ where $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$, for all $\lambda \geq \lambda_2$ that t_{min} is non-decreasing, but now either one or both of λ_1 and λ_2 can be on a transition and there is only one state in between them with no transitions. Again one exception has to be made for the $ILG(j)$ state.

The method of proof is to show that there are always two points α and β, where $\lambda_1 < \alpha < \beta < \lambda_2$, that satisfy the requirements of the first part of the proof.

1. $SRG(j)$ State Derivations

   A list of possible options for λ_1 and λ_2:

   - $T\_SR(j)$ to $SRG(j)$
   - $SRG(j)$ to $T\_SR(j)$
   - $T\_SR(j)$ to $T\_SR(j)$
   - $LOW\_SR(j)$ to $SRG(j)$
   - $LOW\_SR(j)$ to $T\_SR(j)$
   - $SRG(j)$ to $T\_BDY(j, j+1)$
   - $T\_SR(j)$ to $T\_BDY(j, j+1)$
   - $LOW\_SR(j)$ to $T\_BDY(j, j+1)$
   - $T\_BDY(j-1, j)$ to $T\_BDY(j, j+1)$ where $REG(j) = SR$
   - $T\_BDY(j-1, j)$ to $SRG(j)$
   - $T\_BDY(j-1, j)$ to $T\_SR(j)$
   - $HIGH\_BDY(j-1, j)$ to $SRG(j)$

- $HIGH\_BDY(j-1,j)$ to $T\_SR(j)$
- $HIGH\_BDY(j-1,j)$ to $T\_BDY(j,j+1)$ where $REG(j) = SR$

For each of these options $B_{l,1}/C_l \geq 0$ is implied from $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$. Points α and β always exist because state $SRG(j)$ has a none zero range.

2. $ILG(j)$ State Derivations

- $T\_IL(j)$ to $ILG(j)$
- $ILG(j)$ to $T\_IL(j)$
- $T\_IL(j)$ to $T\_IL(j)$
- $LOW\_IL(j)$ to $ILG(j)$
- $LOW\_IL(j)$ to $T\_IL(j)$
- $ILG(j)$ to $T\_BDY(j,j+1)$
- $T\_IL(j)$ to $T\_BDY(j,j+1)$
- $LOW\_IL(j)$ to $T\_BDY(j,j+1)$
- $T\_BDY(j-1,j)$ to $T\_BDY(j,j+1)$ where $REG(j) = IL$
- $T\_BDY(j-1,j)$ to $ILG(j)$
- $T\_BDY(j-1,j)$ to $T\_IL(j)$
- $HIGH\_BDY(j-1,j)$ to $ILG(j)$
- $HIGH\_BDY(j-1,j)$ to $T\_IL(j)$
- $HIGH\_BDY(j-1,j)$ to $T\_BDY(j,j+1)$ where $REG(j) = IL$

For the first three options, an extra requirement is needed, $B_{l,1}/\Delta_{p,j} \geq 0$. For the rest of the options this is already implied.

3. $BDY(j-1,j)$ State Derivations

- $HIGH\_BDY(j-1,j)$ to $BDY(j-1,j)$
- $BDY(j-1,j)$ to $HIGH\_BDY(j-1,j)$
- $HIGH\_BDY(j-1,j)$ to $HIGH\_BDY(j-1,j)$
- $BDY(j-1,j)$ to $TMIN\_CHG$
- $TMIN\_CHG$ to $BDY(j-1,j)$
- $TMIN\_CHG$ to $HIGH\_BDY(j-1,j)$

For $t_{min}(\lambda_1) \leq t_{min}(\lambda_2)$ implies that $B_{k,3} \geq 0$.

The final part is to prove Result 1 from parts 1 and 2. λ_1 and λ_2 are on any transition or state, with any finite number of states and transitions in between. Divide the domain $[\lambda_1, \lambda_2]$ by ζ_i the position of each transition. Now there are a finite number of periods (ζ_{i-1}, ζ_i) each with their own state. Within each period, $dt_{min}/d\lambda$ is constant, as γ_j is always linearly dependent upon λ. So $t_{min}(\lambda)$ is a piecewise function of linear segments.

To prove the result it has to be shown for two cases:

1. If λ_2 is in a state j, that $d\gamma_j/d\lambda > 0$.
2. If λ_2 is on a transition, that for the next state j, $d\gamma_j/d\lambda > 0$.

therefore by the results of part 1 and 2, for any α and β, where $\lambda_2 \leq \alpha \leq \beta \leq 1$, then $t_{min}(\alpha) \leq t_{min}(\beta)$.

For t_{min} to increase from λ_1 to λ_2, two things could have happened:

1. there is a period (ζ_{i-1}, ζ_i) where t_{min} is increasing, then by the results of part 1, both cases are satisfied.
2. or there is a transition, ζ_i, where $t_{min}(\zeta_i^-) < t_{min}(\zeta_i^+)$. There is only one transition where time jumps, T_BDY, and only jumps up in time if the state changes to a higher one, e.g.:

   $SRG(j)$ to $T\_BDY(j, j+1)$ to $SRG(j+1)$

   It has been shown in analysing all the transitions that in the next state $B_{l+1,1} > 0$. Therefore both cases are proven by the results of part 1.

A.6 AREA FOR THE MINIMUM FREQUENCY

In Eq. 5.50 the area function, $A(\mathbf{w})$, is defined. In Eq. A.12, the reserve dispatch by t_{min}, q_i and w_i, are defined. These two sets of equations are combined to find the area function dependent on λ, $A(\lambda)$. The general area contribution from offer i and region j is by the trapezium rule:

$$A_{i,j} = \mu_{i,j}\Delta_{i,j}(\tau_{j,s} + \frac{1}{2}\mu_{i,j}(\tau_{j,e} - \tau_{j,s})) \tag{A.55}$$

If $REG(j) = IL$ then the area is simplified to:

$$A_{i,j} = \mu_{i,j}\Delta_{i,j}\tau_{j,e} \tag{A.56}$$

The total area for each offer is:

$$A_i = \frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}(\tau_{j,s} + \frac{1}{2}\mu_{i,j}(\tau_{j,e} - \tau_{j,s})) \tag{A.57}$$

where t_i is $t_{i,p}$ if the offer is IL, and $t_{i,u}$ if SR. The region j is the one where $0 < \mu_{i,j} < 1$. If on the boundary then pick just the first term with appropriate j.

$$A_i = \frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s})$$

It is also desirable to have the area in terms of λ. The derivation starts by finding the time SR offer stops ramping:

$$t_{i,e} = (1-\lambda)(t_i + \frac{v_{i,1}}{\Delta_i}) + \lambda(t_i + \frac{v_{i,2}}{\Delta_i})$$

$$t_{i,e} = t_i + \frac{1}{\Delta_i}(v_{i,1} + \lambda(v_{i,2} - v_{i,1})) \tag{A.58}$$

where $\Delta_i = u_l^{max}$, i and l refer to the same SR offer, but not necessarily have the same index number. The average time when ramping is:

$$t_{i,h} = \frac{1}{2}(t_i + t_{i,e})$$

$$t_{i,h} = t_i + \frac{1}{2\Delta_i}(v_{i,1} + \lambda(v_{i,2} - v_{i,1})) \tag{A.59}$$

Then the area of each SR offer is:

$$A_i(\lambda) = \Big(t_i + \frac{1}{2\Delta_i}(v_{i,1} + \lambda(v_{i,2} - v_{i,1}))\Big)\Big(v_{i,1} + \lambda(v_{i,2} - v_{i,1})\Big)$$

$$A_i(\lambda) = v_{i,1}(t_i + \frac{v_{i,1}}{2\Delta_i}) + \lambda(v_{i,2} - v_{i,1})(t_i + \frac{v_{i,1}}{\Delta_i}) + \lambda^2\frac{(v_{i,2} - v_{i,1})^2}{2\Delta_i} \tag{A.60}$$

The area for each IL offer is:

$$A_i(\lambda) = t_i v_{i,1} + \lambda t_i(v_{i,2} - v_{i,1}) \tag{A.61}$$

The area $A(\lambda)$ is found for each state:

1. $SRG(j)$ State

$$A(\lambda) = \sum_{\substack{i \in D \\ OFF(i)=IL}} t_i v_{i,1} + \lambda t_i(v_{i,2} - v_{i,1})$$

$$+ \sum_{\substack{i \in D \\ OFF(i)=SR}} \left(v_{i,1}(t_i + \frac{v_{i,1}}{2\Delta_i}) + \lambda(v_{i,2} - v_{i,1})(t_i + \frac{v_{i,1}}{\Delta_i}) + \lambda^2\frac{(v_{i,2} - v_{i,1})^2}{2\Delta_i}\right)$$

$$+\sum_{i\in E}\left(\frac{1}{2}\theta_{i,j,s}(t_i+\tau_{j,s})+\gamma_j\Delta_{i,j}\tau_{j,s}+\frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e}-\tau_{j,s})\right) \tag{A.62}$$

where D and E are the same defined for Eq. A.16, and γ_j is defined in Eq. A.17.

2. $ILG(j)$ State

$$A(\lambda)=\sum_{\substack{i\in D\\ OFF(i)=IL}} t_i v_{i,1}+\lambda t_i(v_{i,2}-v_{i,1})+\sum_{i\in E}\frac{1}{2}\theta_{i,j,s}(t_i+\tau_{j,s})$$

$$+\sum_{\substack{i\in D\\ OFF(i)=SR}}\left(v_{i,1}(t_i+\frac{v_{i,1}}{2\Delta_i})+\lambda(v_{i,2}-v_{i,1})(t_i+\frac{v_{i,1}}{\Delta_i})+\lambda^2\frac{(v_{i,2}-v_{i,1})^2}{2\Delta_i}\right)$$

$$+\gamma_j t_k\Delta_{k,j} \tag{A.63}$$

where D and E are defined in Eq. A.44, and γ_j in Eq. A.45. The IL offer k is the one associated with region j.

3. $BYD(j-1,j)$ State

$$A(\lambda)=\sum_{\substack{i\in D\\ OFF(i)=IL}} t_i w_{i,1}+\lambda t_i(v_{i,2}-v_{i,1})+\sum_{i\in E}\frac{1}{2}\theta_{i,j,s}(t_i+\tau_{j,s})$$

$$+\sum_{\substack{i\in D\\ OFF(i)=SR}}\left(v_{i,1}(t_i+\frac{v_{i,1}}{2\Delta_i})+\lambda(v_{i,2}-v_{i,1})(t_i+\frac{v_{i,1}}{\Delta_i})+\lambda^2\frac{(v_{i,2}-v_{i,1})^2}{2\Delta_i}\right) \tag{A.64}$$

where D and E are defined in Eq. A.44.

A.7 THE SECOND MAIN RESULT - THE MINIMUM FREQUENCY AREA

In order to prove convexity it is necessary to show that the $A(\lambda)$ has a specific shape. The second result is proving $dA/d\lambda$ is always increasing.

Result 2: *The second derivative of $A(\lambda)$ is a non-negative constant, $d^2A/d\lambda^2=\alpha_i\geq 0$ for λ in an open interval $(\lambda_{i-1},\lambda_i)$, and for $\lambda=\lambda_i$:*

$$\left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_T^-}\leq\left.\frac{dA}{d\lambda}\right|_{\lambda=\lambda_T^+} \tag{A.65}$$

The method of proving this result is to show it holds for every state and transition. The general formulas for the first and second derivatives are:

$$\frac{dA}{d\lambda} = \frac{\partial A}{\partial \lambda} + \frac{\partial A}{\partial \gamma_j}\frac{d\gamma_j}{d\lambda} \tag{A.66}$$

$$\frac{d^2A}{d\lambda^2} = \frac{\partial^2 A}{\partial \lambda^2} + 2\frac{\partial^2 A}{\partial \lambda \partial \gamma_j}\frac{d\gamma_j}{d\lambda} + \frac{\partial^2 A}{\partial \gamma_j^2}\left(\frac{d\gamma_j}{d\lambda}\right)^2 + \frac{\partial A}{\partial \gamma_j}\frac{d^2\gamma_j}{d\lambda^2}$$

which simplifies because $\partial^2 A/\partial\lambda\partial\gamma_j = 0$ and $d^2\gamma_j/d\lambda^2 = 0$.

$$\frac{d^2A}{d\lambda^2} = \frac{\partial^2 A}{\partial \lambda^2} + \frac{\partial^2 A}{\partial \gamma_j^2}\left(\frac{d\gamma_j}{d\lambda}\right)^2 \tag{A.67}$$

Proving result 2 for each of the states:

1. $SRG(j)$ State

$$\frac{d^2A}{d\lambda^2} = \left(\sum_{\substack{i\in D \\ OFF(i)=SR}} \frac{(v_{i,2}-v_{i,1})^2}{\Delta_i}\right) + \left(\sum_{i\in E}\Delta_{i,j}(\tau_{j,e}-\tau_{j,s})\right)\left(\frac{B_{l,1}}{C_l}\right)^2 \geq 0 \tag{A.68}$$

2. $ILG(j)$ and $BDY(j-1,j)$ State

$$\frac{d^2A}{d\lambda^2} = \sum_{\substack{i\in D \\ OFF(i)=SR}} \frac{(v_{i,2}-v_{i,1})^2}{\Delta_i} \geq 0 \tag{A.69}$$

For each of the transitions at λ_T there are two separate equations for the area, A_l for $\lambda \leq \lambda_T$ and A_{l+1} for $\lambda \geq \lambda_T$. The general relationship between the two areas is:

$$A_{l+1}(\lambda) = A_l(\lambda) + X_l(\lambda) \tag{A.70}$$

where X_l is the general change that relates the two areas. To prove result 2 is to show $dX_l/d\lambda = 0$ when t_{min} is continuous, and $dX_l/d\lambda \geq 0$ when discontinuous.

1. T_SR and $LOW\_SR(j)$ Transitions

   The area formula changes because there are a set of SR offers, D_T, that transfer from D to E, and also a set E_{AB} that go from E to D.

$$X_l = \sum_{i\in D_T}\Big((\gamma_{j,l+1}-\mu_{i,j})\Delta_{i,j}\tau_{j,s}+\frac{1}{2}(\gamma_{j,l+1}^2-\mu_{i,j}^2)\Delta_{i,j}(\tau_{j,e}-\tau_{j,s})\Big)$$
$$+\sum_{i\in E_{AB}}\Big((\mu_{i,j}-\gamma_{j,l})\Delta_{i,j}\tau_{j,s}+\frac{1}{2}(\mu_{i,j}^2-\gamma_{j,l}^2)\Delta_{i,j}(\tau_{j,e}-\tau_{j,s})\Big)$$
$$+\sum_{i\in E\setminus E_{AB}}\Big((\gamma_{j,l+1}-\gamma_{j,l})\Delta_{i,j}\tau_{j,s}+\frac{1}{2}(\gamma_{j,l+1}^2-\gamma_{j,l}^2)\Delta_{i,j}(\tau_{j,e}-\tau_{j,s})\Big) \quad \text{(A.71)}$$

The derivative is, using the equation for $\mu_{i,j}$ in Eq. A.5:

$$\frac{dX_l}{d\lambda} = \sum_{i\in D_T}\Big(\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{B_{l+1,1}}{C_{l+1}}-\sum_{i\in D_T}\Big(\tau_{j,s}+\mu_{i,j}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{v_{i,2}-v_{i,1}}{\Delta_{i,j}}$$
$$+\sum_{i\in E_{AB}}\Big(\tau_{j,s}+\mu_{i,j}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{v_{i,2}-v_{i,1}}{\Delta_{i,j}}-\sum_{i\in E_{AB}}\Big(\tau_{j,s}+\gamma_{j,l}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{B_{l,1}}{C_l}$$
$$+\sum_{i\in E\setminus E_{AB}}\Big(\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{B_{l+1,1}}{C_{l+1}}-\sum_{i\in E\setminus E_{AB}}\Big(\tau_{j,s}+\gamma_{j,l}(\tau_{j,e}-\tau_{j,s})\Big)\Delta_{i,j}\frac{B_{l,1}}{C_l} \quad \text{(A.72)}$$

At λ_T it is known that $\gamma_{j,l+1}=\gamma_{j,l}=\mu_{i,j}$ for each of the terms in the derivative. Therefore it simplifies to:

$$\frac{dX_l}{d\lambda} = \Big(\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})\Big)\Big(\sum_{i\in D_T}\Big(\Delta_{i,j}\frac{B_{l+1,1}}{C_{l+1}}-(v_{i,2}-v_{i,1})\Big)$$
$$+\sum_{i\in E_{AB}}\Big((v_{i,2}-v_{i,1})-\Delta_{i,j}\frac{B_{l,1}}{C_l}\Big)+\sum_{i\in E\setminus E_{AB}}\Big(\Delta_{i,j}\frac{B_{l+1,1}}{C_{l+1}}-\Delta_{i,j}\frac{B_{l,1}}{C_l}\Big)\Big) \quad \text{(A.73)}$$

$$=\frac{\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})}{C_lC_{l+1}}\Big(\sum_{i\in D_T}\Big(\Delta_{i,j}B_{l+1,1}C_l-C_{l+1}C_l(v_{i,2}-v_{i,1})\Big)$$
$$+\sum_{i\in E_{AB}}\Big(C_{l+1}C_l(v_{i,2}-v_{i,1})-\Delta_{i,j}B_{l,1}C_{l+1}\Big)+\sum_{i\in E\setminus E_{AB}}\Big(\Delta_{i,j}B_{l+1,1}C_l-\Delta_{i,j}B_{l,1}C_{l+1}\Big)\Big) \quad \text{(A.74)}$$

$$=\frac{\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})}{C_lC_{l+1}}\Big(\sum_{\substack{i\in(E\setminus E_{AB})\\ \cup D_T}}\Delta_{i,j}B_{l+1,1}C_l-\sum_{i\in E}\Delta_{i,j}B_{l,1}C_{l+1}$$

$$+C_{l+1}C_l\Big(\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})-\sum_{i\in D_T}(v_{i,2}-v_{i,1})\Big)\Big) \tag{A.75}$$

The following substitutions are made:

$$C_l=\sum_{i\in E}\Delta_{i,j}\quad\text{and}\quad C_{l+1}=\sum_{\substack{i\in(E\setminus E_{AB})\\ \cup D_T}}\Delta_{i,j}$$

$$B_{l+1,1}=B_{l,1}-\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})+\sum_{i\in D_T}(v_{i,2}-v_{i,1}) \tag{A.76}$$

the result is:

$$\frac{dX_l}{d\lambda}=\Big(\tau_{j,s}+\gamma_{j,l+1}(\tau_{j,e}-\tau_{j,s})\Big)\Big(B_{l+1,1}-B_{l,1}+\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})-\sum_{i\in D_T}(v_{i,2}-v_{i,1})\Big)=0 \tag{A.77}$$

2. $T\_IL(j)$ and $LOW\_IL(j)$ Transitions

There are SR offers that transfer from D to E, D_{BA}, and vice versa E_{AB}.

$$\begin{aligned}X_l=&-\sum_{i\in D_{BA}}\left(v_{i,1}(t_i+\frac{v_{i,1}}{2\Delta_i})+\lambda(v_{i,2}-v_{i,1})(t_i+\frac{v_{i,1}}{\Delta_i})+\lambda^2\frac{(v_{i,2}-v_{i,1})^2}{2\Delta_i}\right)\\&+\sum_{i\in D_{BA}}\frac{1}{2}\theta_{i,j,s}(t_i+\tau_{j,s})-\sum_{i\in E_{AB}}\frac{1}{2}\theta_{i,j,s}(t_i+\tau_{j,s})\\&+\sum_{i\in E_{AB}}\left(v_{i,1}(t_i+\frac{v_{i,1}}{2\Delta_i})+\lambda(v_{i,2}-v_{i,1})(t_i+\frac{v_{i,1}}{\Delta_i})+\lambda^2\frac{(v_{i,2}-v_{i,1})^2}{2\Delta_i}\right)\\&+(\gamma_{j,l+1}-\gamma_{j,l})t_k\Delta_{k,j}\end{aligned} \tag{A.78}$$

The derivative is:

$$\begin{aligned}\frac{dX_l}{d\lambda}=&-\sum_{i\in D_{BA}}(v_{i,2}-v_{i,1})\Big(t_i+\frac{v_{i,1}}{\Delta_i}+\lambda\frac{v_{i,2}-v_{i,1}}{\Delta_i}\Big)\\&+\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})\Big(t_i+\frac{v_{i,1}}{\Delta_i}+\lambda\frac{v_{i,2}-v_{i,1}}{\Delta_i}\Big)\\&+t_k\Delta_{k,j}\Big(\frac{B_{l+1,1}}{\Delta_{k,j}}-\frac{B_{l,1}}{\Delta_{k,j}}\Big)\end{aligned} \tag{A.79}$$

At λ_T the time SR offers switch from before or after t_k:

$$\forall i \in D_{BA} \cup E_{AB} \quad t_k = t_i + \frac{v_{i,1} + \lambda_T(v_{i,2} - v_{i,1})}{\Delta_i} \tag{A.80}$$

The derivative becomes:

$$\frac{dX_l}{d\lambda} = t_k\Big(B_{l+1,1} - B_{l,1} - \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1})\Big) = 0 \tag{A.81}$$

The derivative is zero due to Eq. A.47 and A.49.

3. $SRG(j)$ to $HIGH\_BDY(j)$ to $BDY(j-1, j)$

Like for the $SRG(j)$ transition there are two sets D_T and E_{AB} of SR offers transitioning from either side.

$$\begin{aligned} X_l = &- \sum_{i \in D_T} \Big(\frac{1}{2}\theta_{i,j-1,s}(t_i + \tau_{j-1,s}) + \mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s} + \frac{1}{2}\mu_{i,j-1}^2\Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big) \\ &+ \sum_{i \in D_T} \frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) - \sum_{i \in E \setminus E_{AB}} \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \\ &+ \sum_{i \in E_{AB}} \Big(\frac{1}{2}\theta_{i,j-1,s}(t_i + \tau_{j-1,s}) + \mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s} + \frac{1}{2}\mu_{i,j-1}^2\Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big) \\ &- \sum_{i \in E_{AB}} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s})\Big) \end{aligned} \tag{A.82}$$

The above equation assumes $REG(j-1) = SR$, but if $REG(j-1) = IL$ it is matter of replacing $j-1$ with $j-2$, the analysis works for both. The derivative is:

$$\begin{aligned} \frac{dX_l}{d\lambda} = &- \sum_{i \in D_T} \Big(\Delta_{i,j-1}\tau_{j-1,s} + \mu_{i,j-1}\Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big)\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j-1}} \\ &- \sum_{i \in E} \Big(\Delta_{i,j}\tau_{j,s} + \gamma_j\Delta_{i,j}(\tau_{j,e} - \tau_{j,s})\Big)\frac{B_{l,1}}{C_l} \\ &+ \sum_{i \in E_{AB}} \Big(\Delta_{i,j-1}\tau_{j-1,s} + \Delta_{i,j-1}\mu_{i,j-1}(\tau_{j,e} - \tau_{j,s})\Big)\frac{v_{i,2} - v_{i,1}}{\Delta_{i,j-1}} \end{aligned} \tag{A.83}$$

Since the dispatch is on a boundary, $\forall i \in D_T \cup E_{AB} \quad \mu_{i,j-1} = 1$, and $\gamma_j = 0$.

Plus the substitution:

$$C_l = \sum_{i \in E} \Delta_{i,j} \tag{A.84}$$

The derivative simplifies to:

$$\frac{dX_l}{d\lambda} = \tau_{j,s}\Big(- B_{l,1} - \sum_{i \in D_T} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1})\Big) \tag{A.85}$$

substituting:

$$B_{l,1} = (R_2 - R_1) - \sum_{i \in D} (v_{i,2} - v_{i,1}) \tag{A.86}$$

$$\frac{dX_l}{d\lambda} = -\tau_{j,s}\Big((R_2 - R_1) - \sum_{\substack{i \in (D \setminus D_T) \\ \cup E_{AB}}} (v_{i,2} - v_{i,1})\Big) = 0 \tag{A.87}$$

The derivative is equal to zero because of the definition of the $BDY(j-1,j)$ state.

4. $BDY(j-1,j)$ to $HIGH\_BDY(j)$ to $SRG(j)$

   The opposite to the previous transition, define D and E according to the definitions of Eq. A.13 and A.14. The SR offers that change are D_{BA} and E_{BA}. The area between transitions is the opposite of Eq. A.82, so that the derivative becomes:

$$\frac{dX_l}{d\lambda} = \tau_{j,s}\Big(B_{l+1,1} - \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1})\Big) \tag{A.88}$$

   where

$$B_{l+1,1} = (R_2 - R_1) - \sum_{\substack{i \in (D \setminus D_{BA}) \\ \cup E_{AB}}} (v_{i,2} - v_{i,1}) \tag{A.89}$$

   Therefore $dX_l/d\lambda = 0$, as the definition of the $BDY(j-1,j))$ state requires it.

5. $ILG(j)$ to $HIGH\_BDY(j)$ to $BDY(j-1,j)$

   Similar to the previous transition with $SRG(j)$, the SR offers that transition are D_{BA} and E_{AB}. The change in area between the transition is:

$$X_l = -\gamma_j t_k \Delta_{k,j} + \sum_{i \in D_{BA}} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) - \sum_{i \in E_{BA}} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s})\Big)$$

$$- \sum_{i \in D_{BA}} \Big(\frac{1}{2}\theta_{i,j-1,s}(t_i + \tau_{j-1,s}) + \mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s} + \frac{1}{2}\mu_{i,j-1}^2 \Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big)$$

$$+\sum_{i\in E_{AB}}\Big(\frac{1}{2}\theta_{i,j-1,s}(t_i+\tau_{j-1,s})+\mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s}+\frac{1}{2}\mu^2_{i,j-1}\Delta_{i,j-1}(\tau_{j-1,e}-\tau_{j-1,s})\Big) \tag{A.90}$$

Since $\forall i \in D_{BA}\cup E_{AB} \quad \mu_{i,j-1}=1$, and $t_k=\tau_{j-1,e}=\tau_{j,s}$, the derivative is:

$$\frac{dX_l}{d\lambda}=t_k\Big(-B_{l,1}-\sum_{i\in D_{BA}}(v_{i,2}-v_{i,1})+\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})\Big) \tag{A.91}$$

Substituting in:

$$B_{l,1}=(R_2-R_1)-\sum_{i\in D}(v_{i,2}-v_{i,1}) \tag{A.92}$$

The derivative becomes:

$$\frac{dX_l}{d\lambda}=-t_k\Big((R_2-R_1)-\sum_{\substack{i\in(D\setminus D_{BA})\\ \cup E_{AB}}}(v_{i,2}-v_{i,1})\Big)=0 \tag{A.93}$$

The derivative is equal to zero because of the definition of $BDY(j-1,j)$ state.

6. $BDY(j-1,j)$ to $HIGH\_BDY(j)$ to $ILG(j)$

   Opposite to the previous transition. The derivative is:

$$\frac{dX_l}{d\lambda}=t_k\Big(B_{l+1,1}-\sum_{i\in D_{BA}}(v_{i,2}-v_{i,1})+\sum_{i\in E_{AB}}(v_{i,2}-v_{i,1})\Big)=0 \tag{A.94}$$

7. $SRG(j)$ to $T\_BDY(j,j+1)$ to $SRG(j+1)$ Continuous

   All the transitions that are left are based on T_BDY, with different states transitioning in between. For each option there are two different types: continuous and discontinuous; depending on whether SR offers continue across the transition.

   For the current option there are SR offers that surpass the minimum time, D_T, ones that revert back, E_{BA}, and ones that start ramping at the start of the next transition, F_B. If F_B is empty then the analysis is the same as option 1. Therefore assume it is not empty and the area before and after the transition is:

$$X_l=\sum_{\substack{i\in D_T\\ \cup(E\setminus E_{AB})\\ \cup F_B}}\Big(\frac{1}{2}\theta_{i,j+1,s}(t_i+\tau_{j+1,s})+\gamma_{j+1}\Delta_{i,j+1}\tau_{j+1,s}+\frac{1}{2}\gamma^2_{j+1}\Delta_{i,j+1}(\tau_{j+1,e}-\tau_{j+1,s})\Big)$$

$$
\begin{aligned}
&- \sum_{i \in D_T} \left(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \right) \\
&+ \sum_{i \in E_{AB}} \left(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \right) \\
&- \sum_{i \in E} \left(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \right) \qquad \text{(A.95)}
\end{aligned}
$$

Since this is on the transition $\forall i \in D_T \cup E_{AB} \quad \mu_{i,j} = 1$, $\gamma_j = 1$, and $\gamma_{j+1} = 0$. Therefore the derivative is:

$$
\frac{dX_l}{d\lambda} = \tau_{j,e}\left(B_{l+1,1} - \sum_{i \in D_T} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}) - B_{l,1} \right) \qquad \text{(A.96)}
$$

The equation for $B_{l+1,1}$, in order to satisfy the definition of Eq. A.17, is Eq. A.76. Then $dX_l/d\lambda = 0$. Therefore very similar to the analysis of option 1. Also $SRG(j+1)$ to $T\_BDY(j, j+1)$ to $SRG(j)$ for the continuous case will also give the same result.

8. $SRG(j)$ to $T\_BDY(j, j+1)$ to $SRG(j+1)$ Discontinuous

Now there are no SR offers that change ramping before and after the time of the minimum frequency, but there are SR offers F_S that stop ramping at the transition, and again F_B that begin ramping.

$$
\begin{aligned}
X_l = &- \sum_{i \in F_S} \left(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \right) \\
&+ \sum_{i \in F_S} \left(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \right) \\
&+ \sum_{i \in F_B} \left(\gamma_{j+1}\Delta_{i,j+1}\tau_{j+1,s} + \frac{1}{2}\gamma_{j+1}^2\Delta_{i,j+1}(\tau_{j+1,e} - \tau_{j+1,s}) \right) \qquad \text{(A.97)}
\end{aligned}
$$

At the transition, λ_T, $\forall i \in F_S \quad \mu_{i,j} = \gamma_j$. Also $\gamma_{j+1} = 0$. The derivative is:

$$
\frac{dX_l}{d\lambda} = \Big(\tau_{j,s} + \gamma_j(\tau_{j,e} - \tau_{j,s})\Big)\Big(- B_{l,1} + \sum_{i \in F_S} (v_{i,2} - v_{i,1}) \Big) + B_{l+1,1}\tau_{j+1,s} \qquad \text{(A.98)}
$$

The relationship between $B_{l+1,1}$ and $B_{l,1}$ is:

$$
B_{l+1,1} = B_{l,1} - \sum_{i \in F_S} (v_{i,2} - v_{i,1}) \qquad \text{(A.99)}
$$

The derivative simplifies to:

$$\frac{dX_l}{d\lambda} = B_{l+1,1}\Big(\tau_{j+1,s} - \Big(\tau_{j,s} + \gamma_j(\lambda_T)(\tau_{j,e} - \tau_{j,s})\Big)\Big) \geq 0 \tag{A.100}$$

The derivative is greater than or equal to zero because both, $B_{l+1,1} > 0$ for it to transition from j to $j+1$, and the difference in times are greater than or equal to zero.

9. $SRG(j+1)$ to $T\_BDY(j, j+1)$ to $SRG(j)$ Discontinuous

   This is transitioning in the opposite direction than the previous option. The derivative simplifies to:

$$\frac{dX_l}{d\lambda} = -B_{l,1}\Big(\tau_{j+1,s} - \Big(\tau_{j,s} + \gamma_j(\lambda_T)(\tau_{j,e} - \tau_{j,s})\Big)\Big) \geq 0 \tag{A.101}$$

   For it to be transitioning in this direction, $B_{l,1} < 0$, therefore $dX_l/d\lambda \geq 0$. Note that l refers to $j+1$, and $l+1$ to j in this situation. This is opposite to the previous option.

10. $ILG(j)$ to $T\_BDY(j, j+2)$ to $ILG(j+2)$ Discontinuous

    This is when the time of the minimum frequency transitions from the time of one IL offer to another. This is always a discontinuous transition because there is finite difference in time between offers. The region in between can be either SR or NN, but in the SR case it is required that for all i, $\mu_{i,j+1} = 0$ at λ_T. The difference in area is:

$$X_l = \gamma_{j+2} t_{k+2} \Delta_{k+2,j+2} - \gamma_j t_k \Delta_{k,j} + \mu_{k,j} t_k \Delta_{k,j} \tag{A.102}$$

    The derivative is:

$$\frac{dX_l}{d\lambda} = B_{l+1,1} t_{k+2} - t_k\Big(B_{l,1} - (v_{k,2} - v_{k,1})\Big) \tag{A.103}$$

    From the definitions of $B_{l,1}$ and $B_{l+1,1}$: $B_{l+1,1} = B_{l,1} - (v_{k,2} - v_{k,1})$:

$$\frac{dX_l}{d\lambda} = B_{l+1,1}(t_{k+2} - t_k) > 0 \tag{A.104}$$

    because for the transition to occur $B_{l,1} > 0$, and so $B_{l+1,1} > 0$.

11. $ILG(j+2)$ to $T\_BDY(j, j+2)$ to $ILG(j)$ Discontinuous

    Opposite to the previous option, the derivative is:

$$\frac{dX_l}{d\lambda} = -B_{l,1}(t_{k+2} - t_k) > 0 \tag{A.105}$$

Again note that $j+2$ is connected to l, and j to $l+1$, for this option.

12. $SRG(j)$ to $T\_BDY(j, j+1)$ to $ILG(j+1)$ Continuous

SR offers, D_{BA}, stop ramping before the time of the minimum frequency initially, but then stop ramping after. E_{AB} is the reserve. The difference in area between the transitions is:

$$\begin{aligned} X_l = &- \sum_{i \in D_{BA}} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \Big) \\ &+ \sum_{i \in D_{BA}} \Big(\frac{1}{2}\theta_{i,j+1,s}(t_i + \tau_{j+1,s}) \Big) + \sum_{i \in E \setminus E_{AB}} \frac{1}{2}\theta_{i,j+1,s}(t_i + \tau_{j+1,s}) \\ &+ \sum_{i \in E_{AB}} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \Big) \\ &- \sum_{i \in E} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \Big) \\ &+ \gamma_{j+1} t_k \Delta_{k,j+1} \end{aligned} \tag{A.106}$$

At λ_T, $\forall i \in D_{BA} \cup E_{AB} \quad \mu_{i,j} = 1$, $\gamma_j = 1$. Also $\tau_{j,e} = t_k$. The derivative is:

$$\frac{dX_l}{d\lambda} = t_k \Big(B_{l+1,1} - B_{l,1} - \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}) \Big) = 0 \tag{A.107}$$

For the opposite option, $ILG(j+1)$ to $T\_BDY(j, j+1)$ to $SRG(j)$ continuous, the derivative is:

$$\frac{dX_l}{d\lambda} = t_k \Big(B_{l+1,1} - B_{l,1} + \sum_{i \in D_{BA}} (v_{i,2} - v_{i,1}) - \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}) \Big) = 0 \tag{A.108}$$

13. $SRG(j)$ to $T\_BDY(j, j+1)$ to $ILG(j+1)$ Discontinuous

SR offers F_S transition from stop ramping after the minimum time to before, and the state transitions. The difference in area is:

$$\begin{aligned} X_l = &- \sum_{i \in F_S} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \gamma_j\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\gamma_j^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \Big) \\ &+ \sum_{i \in F_S} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s}) + \mu_{i,j}\Delta_{i,j}\tau_{j,s} + \frac{1}{2}\mu_{i,j}^2\Delta_{i,j}(\tau_{j,e} - \tau_{j,s}) \Big) \\ &+ \gamma_{j+1} t_k \Delta_{k,j+1} \end{aligned} \tag{A.109}$$

For all i in F_S, $\mu_{i,j} = \gamma_j$ at λ_T. The derivative is:

$$\frac{dX_l}{d\lambda} = \Big(\tau_{j,s} + \gamma_j(\lambda_T)(\tau_{j,e} - \tau_{j,s})\Big)\Big(- B_{l,1} + \sum_{i \in F_S} (v_{i,2} - v_{i,1})\Big) + t_k B_{l+1,1}$$

$$\frac{dX_l}{d\lambda} = B_{l+1,1}\Big(t_k - \Big(\tau_{j,s} + \gamma_j(\lambda_T)(\tau_{j,e} - \tau_{j,s})\Big)\Big) \geq 0 \tag{A.110}$$

$B_{l+1,1} > 0$ because $B_{l,1} > 0$, which it has to be in order to transition like this. For the opposite direction $ILG(j+1)$ to $T\_BDY(j, j+1)$ to $SRG(j)$, the derivative is:

$$\frac{dX_l}{d\lambda} = -B_{l,1}\Big(t_k - \Big(\tau_{j,s} + \gamma_j(\lambda_T)(\tau_{j,e} - \tau_{j,s})\Big)\Big) \geq 0 \tag{A.111}$$

$B_{l,1} < 0$ for the state to transition to a lower one.

14. $ILG(j)$ to $T\_BDY(j, j+1)$ to $SRG(j+1)$ Continuous

$$\begin{aligned} X_l = &- \sum_{i \in D_T} \Big(\frac{1}{2}\theta_{i,j-1,s}(t_i + \tau_{j-1,s}) + \mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s} + \frac{1}{2}\mu_{i,j-1}^2\Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big) \\ &- \sum_{i \in E} \Big(\frac{1}{2}\theta_{i,j,s}(t_i + \tau_{j,s})\Big) \\ &+ \sum_{i \in E_{AB}} \Big(\frac{1}{2}\theta_{i,j-1,s}(t_i + \tau_{j-1,s}) + \mu_{i,j-1}\Delta_{i,j-1}\tau_{j-1,s} + \frac{1}{2}\mu_{i,j-1}^2\Delta_{i,j-1}(\tau_{j-1,e} - \tau_{j-1,s})\Big) \\ &+ \sum_{\substack{i \in (E \setminus E_{AB}) \\ \cup D_T}} \Big(\frac{1}{2}\theta_{i,j+1,s}(t_i + \tau_{j+1,s}) + \gamma_{j+1}\Delta_{i,j+1}\tau_{j+1,s} + \frac{1}{2}\gamma_{j+1}^2\Delta_{i,j+1}(\tau_{j+1,e} - \tau_{j+1,s})\Big) \\ &(\mu_{j,k} - \gamma_j)t_k\Delta_{j,k} \end{aligned} \tag{A.112}$$

For this transition, $\mu_{i,j-1} = 1$ for i in D_T and E_{AB}. $\gamma_{j+1} = 0$, and $\tau_{j-1,e} = t_k = \tau_{j+1,s}$. The derivative is:

$$\frac{dX_l}{d\lambda} = t_k\Big(- \sum_{i \in D_T} (v_{i,2} - v_{i,1}) + \sum_{i \in E_{AB}} (v_{i,2} - v_{i,1}) + B_{l+1,1} + (v_{k,2} - v_{k,1}) - B_{l,1}\Big) = 0 \tag{A.113}$$

For the opposite, $SRG(j+1)$ to $T\_BDY(j, j+1)$ to $ILG(j)$ continuous, the

derivative is:

$$\frac{dX_l}{d\lambda} = t_k\Big(\sum_{i\in D_T}(v_{i,2}-v_{i,1}) - \sum_{i\in E_{AB}}(v_{i,2}-v_{i,1}) - B_{l,1} - (v_{k,2}-v_{k,1}) + B_{l+1,1}\Big) = 0 \tag{A.114}$$

15. $ILG(j)$ to $T\_BDY(j, j+2)$ to $SRG(j+2)$ Discontinuous

This option there is a set SR offers, F_B, starting to ramp in $SRG(j+2)$ (or $SRG(j+1)$ if there is no difference in time), where $REG(j+1) = NN$. The difference in area is:

$$X_l = (\mu_{k,j} - \gamma_j)t_k\Delta_{k,j}$$

$$+\sum_{i\in F_B}\Big(\frac{1}{2}\theta_{i,j+2,s}(t_i+\tau_{j+2,s}) + \gamma_{j+2}\Delta_{i,j+2}\tau_{j+2,s} + \frac{1}{2}\gamma_{j+2}^2\Delta_{i,j+2}(\tau_{j+2,e}-\tau_{j+2,s})\Big) \tag{A.115}$$

The derivative is:

$$\frac{dX_l}{d\lambda} = t_k((v_{k,2}-v_{k,1}) - B_{l,1}) + \tau_{j+2,s}B_{l+1,1}$$

$$\frac{dX_l}{d\lambda} = B_{l+1,1}(\tau_{j+2,s}-t_k) \geq 0 \tag{A.116}$$

For the opposite transition, $SRG(j+2)$ to $T\_BDY(j, j+2)$ to $ILG(j)$ discontinuous, the derivative is:

$$\frac{dX_l}{d\lambda} = -B_{l,1}(\tau_{j+2,s}-t_k) \geq 0 \tag{A.117}$$

16. $SRG(j)$ to $T\_BDY(j-1, j+1)$ to $SRG(j+1)$

This is the special situation of $SRG(j)$ transitions where $\gamma_j(\lambda_T) = 0$ and $\max\{\mu_{i,j}(\lambda_T)\}_i = 0$. This has the same analysis as $SRG(j)$ to $T\_BDY(j, j+1)$ to $SRG(j+1)$ Discontinuous.

For $SRG(j)$ to $T\_BDY(j-1, j+1)$ to $ILG(j+1)$, the analysis is in $SRG(j)$ to $T\_BDY(j, j+1)$ to $ILG(j+1)$ discontinuous.

Appendix B

DETAILS OF THE REGION TRANSITION ALGORITHM

In Section 5.3.3 a method of finding the global minimum is shown for an example first, and then it was generalised. The general method is to start in one region which has known feasible points, find the minimum for that region, and use the information about where that local minimum is located to determine the next region to check. Repeating this process until eventually the global minimum is found in a region where it does not lie on any internal border. This Appendix provides details to this process, accounting for difficulties that will arise, and the special rules to remove these problems.

The main concern of this Appendix is not proving whether a global minimum exists and that it has specific uniqueness properties, as this can be derived from the convexity properties of the problem, but rather it shows this algorithm will terminate and not get locked in an infinite cycle between regions. From the current description of the algorithm, it seems impossible for this to occur, as the global minimum is found when it does not lie on any internal boundaries between regions. However, there is no restriction that the global minimum cannot lie on an internal boundary, hence step (3b) in the general description of the algorithm is used to terminate it in Section 5.3.4. The difficultly arises when two consecutive regions in the algorithm share the same local minimum; this Appendix develops rules to determine when this local minimum is the global one, or when it is not and which direction has to be taken to find the global minimum.

B.1 DEFINITIONS

Before the algorithm can be fully expressed, a series of definitions is required to describe the conditional statements.

B.1.1 Region Notation

Section 5.3.1 describes the formulation of each region's equations, depending on whether it is of the IL or SR type. The information describing which region the equations are

for is contained in the sets, Q_B, Q_E, $Q_{B,j}$, W_B, W_T, $W_{B,j}$, and $W_{T,j}$. It would be cumbersome to express the transition rules in terms of these sets as firstly there is an indefinite number of them depending on how many frequency constraints there are, but more importantly the transitions rules are more succinctly expressed in changes of state that the IL and SR offers undergo. This section defines a new state based notation to simplify the transition rules and how this state information relates to these sets.

The state information is contained in four arrays:

1. IL_STATE - This array defines whether an IL offer responds before t_{min}, at t_{min}, or after. These states are identified by the capital letters B (Before), C (Current), and A (After) respectively. The array IL_STATE can is indexed by each IL offer, i, to form an expression for example $IL\_STATE(1) = A$ which says IL offer one responds after the time of the minimum frequency. IL_STATE describes the sets Q_B and Q_E:

   - If $IL\_STATE(i) = B$ then $i \in Q_B$ and by extension $i \notin Q_E$.
   - If $IL\_STATE(i) = C$ then $i \in Q_E$ and $i \notin Q_B$.
   - If $IL\_STATE(i) = A$ then $i \notin Q_B$ and $i \notin Q_E$.

2. SR_STATE - Similar to the IL_STATE with slightly different meanings due to SR offers being able to ramp. If SR offer i has a B state then $t_{i,u} < t_{min}$ but its last possible time it can still ramp, $t_{i,e}^{max}$, is up to or before t_{min}. For state C, again $t_{i,u} < t_{min}$ but now it has to be possible for it to continue ramping after t_{min}. Lastly for state A, $t_{i,u} \geq t_{min}$.

3. SR_CROSS - This array expresses the information contained in W_B and W_T. It is very similar to SR_STATE except for one difference those SR offers i which are $SR\_STATE(i) = C$ have two options: the equations can force that SR offer to stop ramping before t_{min}, the B state, or the equations can force it to continue ramping after t_{min}, the S state (Straddle). These definitional rules are:

   - If $SR\_STATE(i) = B$ then $SR\_CROSS(i) = B$.
   - If $SR\_STATE(i) = A$ then $SR\_CROSS(i) = A$.
   - If $SR\_CROSS(i) = B$ then $i \in W_B$ and $i \notin W_T$.
   - If $SR\_CROSS(i) = S$ then $i \in W_T$ and $i \notin W_B$.
   - If $SR\_CROSS(i) = A$ then $i \notin W_B$ and $i \notin W_T$.

4. SR_FNUM - This array defines the time range the SR offer has to stop ramping within. The possible states always starts from 1 and is numbered sequentially up to a maximum depending how many frequency limits it can continue ramping past. This array partly defines $W_{B,j}$ and $W_{T,j}$.

- If $SR\_CROSS(i) = B$ then always $SR\_FNUM(i) = 1$ as the SR offer can only stop ramping from the time it starts ramping until the time of its maximum or t_{min} whichever one comes earliest. Also it does not matter if frequency constraints occur within this time as they are known to be satisfied already.
- If $SR\_CROSS(i) = S$ and $SR\_FNUM(i) = x$ then the SR offer can only stop ramping after the $x - 1$ frequency constraint after t_{min} and before the x one. If $x = 1$ then the time of t_{min} determines the lower limit, $u_i \geq g_i(t_{min} - t_{i,u})$, and if x has reached its maximum possible value then the higher limit is set by $u_i \leq u_i^{max}$. Otherwise in general $g_i(t_{j-1,l} - t_{i,u}) \leq u_i \leq g_i(t_{j,l} - t_{i,u})$. (In certain situations x may correlate to j the frequency constraint index so that $x = j$, but generally it does not.)
- If $SR\_CROSS(i) = A$ and $SR\_FNUM(i) = x$ then the rules are the same as for $SR\_CROSS(i) = S$, except when $x = 1$ the lower limit is $u_i \geq 0$.

So far most of the information contained in the sets is found in the states except for one, $Q_{B,j}$. This information is omitted from the four arrays because it does not make a distinction between regions, or directly feature in the transition rules as something to change. This information is still kept when required to form the equation set.

B.1.2 Region Index

It is convenient to have a unique identifier for each region to reference them. Each region is identified by a set of three numbers to capture the hierarchy of regions. The three numbers are caused by divisions in time, and major and minor divisions in SR dispatch range. The later two do not feature in the transition rules of this Appendix, but are useful in programming the algorithm. These three numbers combine to form the triple, (T,J,M).

- Time Level - Referenced by the index T, this index captures the range t_{min} is allowed to vary. For each T the set of regions could be IL or SR type. The further t_{min} is allowed to be delayed the higher T rises.
- Major Level - Referenced by the index J, this index determines the option for SR_CROSS within the confines of SR_STATE. If there are N SR offers where $SR\_STATE(i) = C$, and index T is for a IL type region then there are 2^N options, for a SR type region there is one less. Therefore index J can range from 1 to 2^N.
- Minor Level - Referenced by the index M, this index determines the option for SR_FNUM within the confines of SR_STATE and SR_CROSS. The number

Table B.1 State and index information for the example captured in Table 5.1.

Label	T	J	M	IL_STATE	SR_STATE	SR_CROSS	SR_FNUM
(g)	1	1	1	[C]	[C,C]	[B,B]	[1,1]
(h)	1	2	1	[C]	[C,C]	[S,B]	[1,1]
(i)	1	2	2	[C]	[C,C]	[S,B]	[2,1]
(d)	1	3	1	[C]	[C,C]	[B,S]	[1,1]
(a)	1	3	2	[C]	[C,C]	[B,S]	[1,2]
(e)	1	4	1	[C]	[C,C]	[S,S]	[1,1]
(f)	1	4	2	[C]	[C,C]	[S,S]	[2,1]
(b)	1	4	3	[C]	[C,C]	[S,S]	[1,2]
(c)	1	4	4	[C]	[C,C]	[S,S]	[2,2]
(p)	2	1	1	[B]	[C,C]	[S,B]	[1,1]
(q)	2	1	2	[B]	[C,C]	[S,B]	[2,1]
(m)	2	2	1	[B]	[C,C]	[B,S]	[1,1]
(j)	2	2	2	[B]	[C,C]	[B,S]	[1,2]
(n)	2	3	1	[B]	[C,C]	[S,S]	[1,1]
(o)	2	3	2	[B]	[C,C]	[S,S]	[2,1]
(k)	2	3	3	[B]	[C,C]	[S,S]	[1,2]
(l)	2	3	4	[B]	[C,C]	[S,S]	[2,2]
(t)	3	1	1	[B]	[C,C]	[S,B]	[1,1]
(r)	3	2	1	[B]	[C,C]	[B,S]	[1,1]
(s)	3	3	1	[B]	[C,C]	[S,S]	[1,1]
(u)	4	1	1	[B]	[C,B]	[S,B]	[1,1]

of possible options is dependent on how many frequency constraints can occur while SR offers are still ramping.

In problems that do not have any SR offers, the major and minor levels lose their meaning. Whenever a triple is expressed for these regions the J and M indices are always one. In the example defined in Table 5.1, the corresponding state and index information is presented in Table B.1 as an example.

B.1.3 Special Region Types

In Section 5.3.1, the formulations for SR and IL type regions are created. For each type there is a special case that requires a special formulation, call these the SR0 and IL0 type regions respectively. These special cases occur when both Q_B and W_B are empty sets. This can only happen when $T = 1$. The impact this has on the formulation is that both t_{min} or p_m are constants, and the equation set is reduced:

- IL Type

  The known value for p_m is calculated through Eq. 5.81, then that equation is removed from the formulation. If the region has any feasible points, Eq. 5.82 will be satisfied, and is redundant and removed from the formulation. Lastly, the left

hand part of Eq. 5.84 is removed, and p_m is counted as constant on the right hand part.

- SR Type

  For this situation $t_{min} = t_{min,0}$, Eq. 5.88 is the same formula used to calculate $t_{min,0}$ and is removed from the formulation. Eq. 5.89 and 5.91 are removed, and if w_i^{min} and w_i^{max} in Eq. 5.93 depend on t_{min} then it is counted as a constant.

In the transition rules it will be helpful to refer to the region type. Create a function $R\_TYPE(T)$ to retrieve this information, e.g. $R\_TYPE(T) = IL$ is true when for time index T it is a IL type region. Include a second function to express if a region is of the special case, $TYPE\_SPEC(T, J)$. The range for this function is $\{TRUE, FALSE\}$, it is true when it is the special situation.

B.1.4 Boundaries between Minimum Times

The border between regions with different t_{min} ranges, i.e. with different time indices T, is due to four possible scenarios: a frequency constraint, $t_{j,l}$; when IL is initiated, $t_{i,p}$; when SR starts ramping up, $t_{i,u}$; and the last possible time SR can stop ramping, $t_{i,e}^{max}$; or any combination of these scenarios. It is necessary to describe these scenarios, to do this seven functions are created, where each has a set as the output. The inputs are two time indices, T_1 and T_2, where $T_1 < T_2$. It is not necessary for $T_1 + 1 = T_2$, as the functions are generalised for longer separations. For T_1 define t_b^{min} as the maximum extent t_{min} can reach in a T_1 region: for a SR type region this is $t_b^{min} = t_{min}^{max}$, for an IL type region this is $t_b^{min} = t_{min}$. For T_2 define t_b^{max} as the minimum extent t_{min} can reach in a T_2 region. The seven functions are:

- $B\_FCON(T_1, T_2)$ - The result is a set of indices of all the frequency constraints that occur between the two times, $j \in B\_FCON(T_1, T_2)$ if and only if $t_b^{min} \leq t_{j,l} \leq t_b^{max}$.

- $B\_ILLEFT(T_1, T_2)$ - If and only if $R\_TYPE(T_1) = IL$ and $IL\_STATE(i, T_1) = C$ then $i \in B\_ILLEFT(T_1, T_2)$.

- $B\_ILRIGHT(T_1, T_2)$ - If and only if $R\_TYPE(T_2) = IL$ and $IL\_STATE(i, T_2) = C$ then $i \in B\_ILRIGHT(T_1, T_2)$.

- $B\_ILGAP(T_1, T_2)$ - This function returns all the IL initiated between the two regions, $i \in B\_ILGAP(T_1, T_2)$ if and only if $t_b^{min} < t_{i,p} < t_b^{max}$.

- $B\_SRFINISH(T_1, T_2)$ - An SR offer, i, is in this set if and only if $t_b^{min} \leq t_{i,e}^{max} \leq t_b^{max}$ and $t_{i,u} < t_b^{min}$.

- $B\_SRSTART(T_1, T_2)$ - An SR offer, i, is in this set if and only if $t_b^{min} \leq t_{i,u} \leq t_b^{max}$ and $t_{i,e}^{max} > t_b^{max}$.
- $B\_SRGAP(T_1, T_2)$ - This is all the SR offers that can start and finish ramping (at its maximum possible extent) in between the two time based regions. $i \in B\_SRGAP(T_1, T_2)$ if and only if $t_b^{min} \leq t_{i,u} < t_{i,e}^{max} \leq t_b^{max}$.

B.1.5 Status of Inequality Constraints

A function is required to express whether an inequality constraint is binding or not for a local minimum, and the type of constraint it is in the case of the SR limits. The main constraints of interest are the ones that determine if the local minimum are on an internal boundary or not. The function returns a value of $TRUE$ if the constraint is binding, $FALSE$ if it is not. The functions are:

- $TIME\_LOW$ - For a SR type region, this function returns $TRUE$ when the local minimum has $t_{min} = t_{min}^{min}$. For an IL type region it is $TRUE$ when $p_m = 0$.
- $TIME\_HIGH$ - For a SR type region, this function returns $TRUE$ when for the local minimum $t_{min} = t_{min}^{max}$, and for an IL type region, when $s_m = 0$.
- $IL\_LOW(i)$ - It is true when $p_i = 0$.
- $IL\_HIGH(i)$ - It is true when $p_i = p_i^{max}$.
- $SR\_LOW(i)$ - It returns true when either $u_i = 0$ or $s_{i,u}^{min} = 0$, i.e. the slack variable from Eq. 5.146 and 5.161.
- $SR\_HIGH(i)$ - It is true when $s_{i,u}^{max} = 0$, i.e. the corresponding slack variable from Eq. 5.146 and 5.161.

The functions which return the type of SR limit it is are $SR\_TYPE\_LOW(i)$ and $SR\_TYPE\_HIGH(i)$, they return one of three different options:

- $OUTSIDE$ - When the limits are the outside minimum and maximum limits, $u_i \geq 0$, or $u_i \leq u_i^{max}$.
- $FREQCON$ - When the limits are of the form $u_i \leq g_i(t_{j,l} - t_{i,u})$ or $u_i \geq g_i(t_{j,l} - t_{i,u})$.
- $TMINEX$ - When the limits are of the form $u_i \leq g_i(t_{min} - t_{i,u})$ or $u_i \geq g_i(t_{min} - t_{i,u})$.

The information contained in these two functions is found in SR_CROSS and SR_FNUM already.

B.2 ALGORITHM OVERVIEW

The goal of this section is to repeat the general description of the algorithm presented in Section 5.3.4, but in terms of the definitions made in the previous section of this Appendix. This will provide the main structure of the algorithm before detailing the transition rules.

The progress of the algorithm is marked by the index, l. Each time a new region is optimised, l is incremented one. The result from the optimisation is the local minimum, $\mathbf{v}_l = [p_{1,l}, \cdots, p_{Np,l}, u_{1,l}, \cdots, u_{Nu,l}]^T$. In some situations the local minimum may not be unique, therefore choose the local minimum $\mathbf{v}_l$ to be one where the least number of inequality constraints are binding. There may be situations where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ of Eqs. 5.165 to 5.168 are not unique, but this is inconsequential for the algorithm, only recognise when $\mathbf{y}$ is not unique. The function, $UNQ(l)$, is used to return $TRUE$ when the solution is unique, and $FALSE$ when it is not.

The algorithm is as follows:

Main Algorithm

1. Find the starting region. $T = 1$. Choose SR_CROSS where for all SR offers if $SR\_STATE(i) = C$ then $SR\_CROSS(i) = S$. Also choose SR_FNUM where for all i, $SR\_FNUM(i)$ is at its maximum index value.

   This is equivalent to finding the region where all the reserve can be dispatched at its maximum. It is assumed that the feasibility of the problem has already been checked and is feasible. Therefore it is not necessary to check the feasibility of any subsequent regions.

2. $GLOBAL\_NOT\_FOUND \leftarrow TRUE, \quad l \leftarrow 0$

3. **WHILE** $GLOBAL\_NOT\_FOUND$

4. $\quad l \leftarrow l + 1$

5. $\quad$ Find the local minimum. (Section 5.3.5)

6. $\quad$ **IF** $ON\_INTERNAL\_BOUNDARY$

7. $\quad\quad$ **IF** $\Big(UNQ(l) \quad \text{and} \quad UNQ(l-1) \quad \text{and} \quad \mathbf{v}_l = \mathbf{v}_{l-1}\Big)$
   $\quad\quad$ or $\Big((\neg UNQ(l) \quad \text{or} \quad \neg UNQ(l-1)) \quad \text{and} \quad \{\mathbf{v}_l\} \cap \{\mathbf{v}_{l-1}\} \neq \emptyset\Big)$

8. $\quad\quad\quad$ **IF** $IS\_GLOBAL(\mathbf{v}_{l-1}, \mathbf{v}_l)$ (Section B.4)

9. $\quad\quad\quad\quad GLOBAL\_NOT\_FOUND \leftarrow FALSE$

   $\quad\quad\quad$ **ELSE**

10. Find next region. (Section B.5)

END

ELSE

11. Find next region. (Section B.3)

END

ELSE

12. $GLOBAL\_NOT\_FOUND \leftarrow FALSE$

END

END

The algorithm repeatedly finds the local minimum of each region it has to check. If the global minimum has been found then $GLOBAL\_NOT\_FOUND$ is $FALSE$ and the algorithm terminates.

B.2.1 Internal Boundaries

On Line 6 of Main Algorithm, a control variable is introduced $ON\_INTERNAL\_BOUNDARY$, which is $TRUE$ when the local minimum lies on an internal boundary. The conditions for this to be $TRUE$ requires for one of the following statements to hold:

1. $TIME\_LOW \quad \wedge \quad T > 1$
2. $TIME\_HIGH \quad \wedge \quad T < T_{max} \quad \wedge \quad R\_TYPE(T) = SR$
3. $TIME\_HIGH \quad \wedge \quad T < T_{max} \quad \wedge \quad R\_TYPE(T) = IL \quad \wedge \quad \Big(\exists i \quad SR\_CROSS(i) = S \quad \vee \quad \exists i \quad SR\_STATE(i) = A \quad \vee \quad \exists i \quad IL\_STATE(i) = A\Big)$
4. $\exists i \quad \Big(SR\_HIGH(i) \quad \wedge \quad SR\_TYPE\_HIGH(i) \neq OUTSIDE\Big)$
5. $R\_TYPE(T) = IL \quad \wedge \quad \exists i \quad \Big(SR\_LOW(i) \quad \wedge \quad SR\_TYPE\_LOW(i) \neq OUTSIDE\Big)$
6. $R\_TYPE(T) = SR \quad \wedge \quad \exists i \quad \Big(SR\_LOW(i) \quad \wedge \quad SR\_TYPE\_LOW(i) \neq OUTSIDE\Big) \quad \wedge \quad \Big(\exists k \neq i \quad SR\_CROSS(k) = S \quad \vee \quad \exists k \quad SR\_STATE(k) = A \quad \vee \quad \exists k \quad IL\_STATE(k) = A\Big)$

where the operators $\wedge$ and $\vee$ are the logical operators for 'and' and 'or' respectively. In later transition rules, $\neg$ is the logical 'not' operator, and $\emptyset$ is the 'Empty Set'. The conditions are derived from the distinction made between internal and external boundaries made in Section 5.3.4. In the transition rules it is necessary to recognise whether a boundary is internal or not, therefore create the following variables that are $TRUE$ when corresponding conditions above are held:

- I_TIME_LOW: Condition 1.
- I_TIME_HIGH: Condition 2 or 3.
- $I\_SR\_HIGH(i)$: If for SR offer i Condition 4 is held.
- $I\_SR\_LOW(i)$: If for SR offer i either Condition 5 or 6 are held.

B.3 TRANSITION RULES

Transitions rules are required when the local minimum is on an internal boundary, the algorithm requires a new region to find the local minimum for the next step, Line 11 of Main Algorithm. These rules are divided into three cases, the occurrence of the first two cases in any practical problem is uncommon, most transitions are the result of Case 3.

B.3.1 Case 1

Case 1 is the result of there being no clear SR offer, i, that should be $SR\_CROSS(i) = S$ in the next region. For this to be true one of the following conditions has to hold:

1. $R\_TYPE(T) = SR \quad \wedge \quad \nexists i \quad \Big(SR\_CROSS(i) = B \quad \wedge \quad I\_SR\_HIGH(i)\Big) \quad \wedge$ $\forall i \quad \text{where} \quad SR\_CROSS(i) = S \quad \text{then} \quad \Big(SR\_TYPE\_LOW(i) = TMINEX \quad \wedge$ $I\_SR\_LOW(i)\Big)$

2. $R\_TYPE(T) = IL \quad \wedge \quad Nu > 0 \quad \wedge \quad \nexists i \quad SR\_CROSS(i) = S \quad \wedge \quad I\_TIME\_HIGH \quad \wedge$ $\nexists i \quad \Big(SR\_CROSS(i) = B \quad \wedge \quad I\_SR\_HIGH(i)\Big)$

3. $R\_TYPE(T) = IL \quad \wedge \quad Nu > 0 \quad \wedge \quad I\_TIME\_HIGH \quad \wedge$ $\nexists i \quad \Big(SR\_CROSS(i) = B \quad \wedge \quad I\_SR\_HIGH(i)\Big) \quad \wedge \quad \forall i \quad \text{where} \quad SR\_CROSS(i) =$ $S \quad \text{then} \quad \Big(SR\_TYPE\_LOW(i) = TMINEX \quad \wedge \quad I\_SR\_LOW(i)\Big)$

4. $R\_TYPE(T) = IL \quad \wedge \quad Nu = 0 \quad \wedge \quad I\_TIME\_HIGH$

The choice of next region depends on whether there is any reserve dispatched after t_{min} for the local minimum, $\mathbf{v}_l$.

IF

$$\sum_{\substack{i \\ IL\_STATE(i)=A}} p_{i,l} \quad + \quad \sum_{\substack{i \\ SR\_STATE(i)=A}} u_{i,l} \quad > \quad 0$$

THEN

Find the smallest value for T_2 such that:

$$\sum_{i \in B\_ILGAP(T,T_2)} p_{i,l} \quad + \quad \sum_{i \in B\_SRGAP(T,T_2)} u_{i,l} \quad = \quad 0$$

and

$$\sum_{i \in B\_ILRIGHT(T,T_2)} p_{i,l} \quad + \quad \sum_{i \in B\_SRSTART(T,T_2)} u_{i,l} \quad > \quad 0$$

make the following assignments to get the next region:
$\forall i$ such that $IL\_STATE(i) = C: \quad IL\_STATE(i) \leftarrow B$
$\forall i \in B\_ILGAP(T, T_2): \quad IL\_STATE(i) \leftarrow B$
$\forall i \in B\_ILRIGHT(T, T_2): \quad IL\_STATE(i) \leftarrow C$
$\forall i$ such that $SR\_CROSS(i) = S: \quad SR\_CROSS(i) \leftarrow B$
$\forall i \in B\_SRFINSIH(T, T_2): \quad SR\_STATE(i) \leftarrow B$
$\forall i \in B\_SRGAP(T, T_2): \quad SR\_STATE(i) \leftarrow B \quad$ and $\quad SR\_CROSS(i) \leftarrow B$
and $SR\_FNUM(i) \leftarrow 1$
$\forall i \quad$ where $\quad SR\_STATE(i) = A \quad$ and $\quad I\_SR\_HIGH(i):$
$SR\_FNUM(i) \leftarrow SR\_FNUM(i) + 1$
$\forall i \quad$ where $\quad SR\_STATE(i) = A \quad$ and $\quad I\_SR\_LOW(i):$
$SR\_FNUM(i) \leftarrow SR\_FNUM(i) - 1$
$\forall i \in B\_SRSTART(T, T_2) \quad$ and $\quad u_{i,l} = 0: \; SR\_STATE(i) \leftarrow C$
and $SR\_CROSS(i) \leftarrow B$
$\forall i \in B\_SRSTART(T, T_2) \quad$ and $\quad u_{i,l} > 0: \; SR\_STATE(i) \leftarrow C$
and $SR\_CROSS(i) \leftarrow S$
$T \leftarrow T_2$

ELSE

IF $\quad \exists i \quad IL\_STATE(i) = A \quad \vee \quad \exists i \quad SR\_STATE(i) = A$

THEN Find the smallest value for T_2 such that $B\_ILRIGHT(T, T_2)$ is not empty or $B\_SRSTART(T, T_2)$ is not empty. Then the next region is found by making the following assignments:

$\forall i$ such that $IL\_STATE(i) = C: \quad IL\_STATE(i) \leftarrow B$

$\forall i$ such that $SR\_CROSS(i) = S$: $\quad SR\_CROSS(i) \leftarrow B$
$\forall i \in B\_SRFINSIH(T, T_2)$: $\quad SR\_STATE(i) \leftarrow B$
$\forall i \in B\_SRSTART(T, T_2)$: $\quad SR\_STATE(i) \leftarrow C$
and $SR\_CROSS(i) \leftarrow S$
$\forall i \in B\_ILRIGHT(T, T_2)$: $\quad IL\_STATE(i) \leftarrow C$
$T \leftarrow T_2$

ELSE

IF $I\_TIME\_LOW \quad \wedge \quad \neg I\_TIME\_HIGH$

THEN

IF There does not exist an SR offer, i, where $i \notin B\_SRSTART(T-1, T)$ and $SR\_CROSS(i) = S$ or where $i \in B\_SRFINISH(T-1, T)$ and $u_{i,l} = u_i^{max}$ and $R\_TYPE(T-1) = SR$.

THEN Find the maximum T_1 such that there exists a IL_STATE, SR_STATE, and SR_CROSS where

$$\sum_{\substack{i \\ IL\_STATE(i,T_1)=C}} p_{i,l} \quad + \quad \sum_{\substack{i \\ SR\_CROSS(i,T_1)=S}} u_{i,l} \quad > 0$$

IF $R\_TYPE(T_1) = IL$

THEN Make the following assignments:
$\forall i \in B\_SRSTART(T_1, T)$: $\quad SR\_STATE(i) \leftarrow A$
and $SR\_CROSS(i) \leftarrow A$
$\forall i \in B\_SRGAP(T_1, T)$: $\quad SR\_STATE(i) \leftarrow A$
and $SR\_CROSS(i) \leftarrow A$
$\forall i \in B\_ILRIGHT(T_1, T)$: $\quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_ILGAP(T_1, T)$: $\quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_ILLEFT(T_1, T)$: $\quad IL\_STATE(i) \leftarrow C$
$T \leftarrow T_1$

ELSE $R\_TYPE(T_1) = SR$
Find the SR offer, i, where $SR\_CROSS(i) = B$, that finished ramping last, i.e. find the maximum $t_{i,u} + u_{i,l}/g_i$. If there is one or more instances of the maximum value, then apply the following to each: $SR\_CROSS(i) \leftarrow S$.

$\forall i \in B\_SRFINISH(T_1, T): \quad SR\_STATE(i) \leftarrow C$
$\forall i \in B\_SRSTART(T_1, T): \quad SR\_STATE(i) \leftarrow A$
and $SR\_CROSS(i) \leftarrow A$
$\forall i \in B\_SRGAP(T_1, T): \quad SR\_STATE(i) \leftarrow A$
and $SR\_CROSS(i) \leftarrow A$
$\forall i \in B\_ILRIGHT(T_1, T): \quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_ILGAP(T_1, T): \quad IL\_STATE(i) \leftarrow A$
$T \leftarrow T_1$

END

ELSE There does exist one or multiple SR offer, i, where
$i \notin B\_SRSTART(T-1, T)$ and $SR\_CROSS(i) = S$
or where $i \in B\_SRFINISH(T-1, T)$ and $u_{i,l} = u_i^{max}$
and $R\_TYPE(T-1) = SR$. If there are multiple instances,
choose the first one with the lowest price, $c_{i,u}$.

THEN For that one SR offer, i, make the assignments:
$SR\_CROSS(i) \leftarrow S$
$\forall k \neq i$ where $SR\_CROSS(k) = S: \quad SR\_CROSS(k) \leftarrow B$

Make the following general assignments:
$\forall i \in B\_SRSTART(T-1, T): \quad SR\_STATE(i) \leftarrow A$
and $SR\_CROSS(i) \leftarrow A$

IF $R\_TYPE(T-1) = IL$ and $\sum_{i \in B\_ILLEFT(T-1,T)} p_{i,l} = 0$
THEN $\forall i \in B\_ILLEFT(T-1, T): \quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_SRFINISH(T-1, T): \quad SR\_STATE(i) \leftarrow C$
$T \leftarrow T - 2$

ELSEIF $R\_TYPE(T-1) = IL$
THEN $\forall i \in B\_ILLEFT(T-1, T): \quad IL\_STATE(i) \leftarrow C$
$T \leftarrow T - 1$

ELSE $R\_TYPE(T-1) = SR$
$\forall i \in B\_SRFINISH(T-1, T): \quad SR\_STATE(i) \leftarrow C$
$T \leftarrow T - 1$

END

END

ELSEIF I_TIME_HIGH
THEN Find the SR offer, i, where $SR\_CROSS(i) = S$ and $i \notin B\_SRFINISH(T, T+1)$,

with the lowest price, $c_{i,u}$. For that one SR offer make the following assignments: $\forall k \neq i$ where $SR\_CROSS(k) = S$: $\quad SR\_CROSS(k) \leftarrow B$

Make the following general assignments:
$\forall i \in B\_SRFINISH(T, T+1)$: $\quad SR\_STATE(i) \leftarrow B$
$\forall i \in B\_ILLEFT(T, T+1)$: $\quad IL\_STATE(i) \leftarrow B$
$T \leftarrow T + 1$

ELSE

Find the SR offer, i, where $SR\_CROSS(i) = S$ with the lowest price, $c_{i,u}$.
For that SR offer, i: $\forall k \neq i$ where $SR\_CROSS(k) = S$:
$SR\_CROSS(k) \leftarrow B$

END

END

END

B.3.2 Case 2

The second case is similar to the first when there is no clear SR offer, i, that should be $SR\_CROSS(i) = S$ in the next region. However the type of boundary the local minimum is on in this case is different. Starting with none of the four conditions for Case 1 being true, the conditions for case two are:

1. $R\_TYPE(T) = IL \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad \nexists i \quad SR\_CROSS(i) = S \quad \wedge$ $\forall i$ where $SR\_STATE(i) = C$ and $SR\_CROSS(i) = B$ then $\neg I\_SR\_HIGH(i)$
2. $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad \forall i$ where $SR\_CROSS(i) = S$ then $i \in B\_SRSTART(T-1, T)$

Finding the next region is dependent on finding the last reserve to be dispatched before the start of a region with index T. Find the maximum T_1 such that there exists a IL_STATE, SR_STATE, and SR_CROSS where:

$$\sum_{\substack{i \\ IL\_STATE(i,T_1)=C}} p_{i,l} \quad + \quad \sum_{\substack{i \\ SR\_CROSS(i,T_1)=S}} u_{i,l} \quad > 0$$

The first assignments are made to the original state arrays:
$\forall i \in B\_ILGAP(T1, T)$: $\quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_SRGAP(T1, T)$: $\quad SR\_STATE(i) \leftarrow A$ and $SR\_CROSS(i) \leftarrow A$
$\forall i \in B\_ILRIGHT(T_1, T)$: $\quad IL\_STATE(i) \leftarrow A$
$\forall i \in B\_SRSTART(T_1, T)$: $\quad SR\_STATE(i) \leftarrow A$ and $SR\_CROSS(i) \leftarrow A$

The next depends on the type of region T_1 is:

IF $R\_TYPE(T_1) = IL$

THEN Make the following assignments:

$\forall i \in B\_ILLEFT(T_1, T): \quad IL\_STATE(i) \leftarrow C$

$T \leftarrow T_1$

ELSE $R\_TYPE(T_1) = SR$

THEN Find the SR offer, i, where $SR\_STATE(i) = C$, with the latest time it has stopped ramping, $t_{i,u} + u_{i,l}/g_i$. If there is one or multiple instances, then apply the following to each:

$SR\_CROSS(i) \leftarrow S$

Make the following general assignments:

$\forall i \in B\_SRFINISH(T_1, T): \quad SR\_STATE(i) \leftarrow C$

$T \leftarrow T_1$

END

B.3.3 Case 3

The third case is for situations where there is no ambiguity for which SR offer will have the state $SR\_CROSS(i) = S$. Case 3 applies to all situations not covered by the conditions of the first two cases. The responses for this case are listed in 'if' and 'then' statements, which try to avoid nesting. A distinction needs to be made between the state arrays from one step to another, these assignments are made from the start:

$IL\_STATE(i, l+1) \leftarrow IL\_STATE(i, l)$

$SR\_STATE(i, l+1) \leftarrow SR\_STATE(i, l)$

$SR\_CROSS(i, l+1) \leftarrow SR\_CROSS(i, l)$

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l)$

These are the series of conditional statements:

IF $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad R\_TYPE(T-1) = SR$
$\wedge \quad B\_FCON(T-1, T) \neq \emptyset$

(Label the following instructions $RULES\_SRDOWNSR\_FREQ$)

THEN $\forall i \in B\_SRFINISH(T-1, T)$ where $u_{i,l} = u_i^{max}: \quad SR\_CROSS(i) \leftarrow S$

$\forall i \in B\_SRFINISH(T-1, T): \quad SR\_STATE(i, l+1) \leftarrow C$

$\forall i \in B\_SRSTART(T-1, T): \quad SR\_STATE(i, l+1) \leftarrow A$ and $SR\_CROSS(i, l+1) \leftarrow A$

$\forall i$ where $SR\_CROSS(i, l+1) = A$, $SR\_FNUM(i, l) \geq 2$, and $I\_SR\_LOW(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$

$\forall i$ where $SR\_CROSS(i, l+1) = A$, $SR\_CROSS(i, l) \neq B$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$\forall i$ where $SR\_CROSS(i, l) = S$ and $i \notin B\_SRSTART(T-1, T)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRSTART(T-1, T)$, and $I\_SR\_LOW(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRSTART(T-1, T)$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$\forall i$ where $SR\_CROSS(i, l) = B$, $i \notin B\_SRSTART(T-1, T)$, and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i, l) \leftarrow S$ and $SR\_FNUM(i, l+1) \leftarrow 2$
$T \leftarrow T - 1$

END

IF $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad R\_TYPE(T-1) = SR$
$\wedge \quad B\_FCON(T-1, T) = \emptyset$

$(RULES\_SRDOWNSR\_NOFREQ)$
THEN $\forall i \in B\_SRFINISH(T-1, T)$ where $u_{i,l} = u_i^{max}$: $\quad SR\_CROSS(i) \leftarrow S$
$\forall i \in B\_SRFINISH(T-1, T)$: $\quad SR\_STATE(i, l+1) \leftarrow C$
$\forall i \in B\_SRSTART(T-1, T)$: $\quad SR\_STATE(i, l+1) \leftarrow A$ and $SR\_CROSS(i, l+1) \leftarrow A$
$\forall i$ where $SR\_CROSS(i, l+1) = A$, $SR\_FNUM(i, l) \geq 2$, and $I\_SR\_LOW(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i, l+1) = A$, $SR\_CROSS(i, l) \neq B$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$\forall i$ where $SR\_CROSS(i, l) = B$, $i \notin B\_SRSTART(T-1, T)$, and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i, l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRSTART(T-1, T)$, $SR\_FNUM(i, l) = 1$,
and $I\_SR\_LOW(i)$: $\quad SR\_CROSS(i, l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRSTART(T-1, T)$, $SR\_FNUM(i, l) \geq 2$,
and $I\_SR\_LOW(i)$: $\quad SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRSTART(T-1, T)$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$T \leftarrow T - 1$

END

IL $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad R\_TYPE(T-1) = IL$

THEN $\forall i \in B\_ILLEFT(T-1, T)$: $\quad IL\_STATE(i, l+1) \leftarrow C$
Find the total IL dispatched at this boundary:

$$IL\_TOT = \sum_{\substack{i \\ IL\_STATE(i,l+1)=C}} p_{i,l}$$

IF $IL\_TOT > 0 \quad \vee \quad T-1 = 1$

THEN $\forall i \in B\_SRSTART(T-1,T): \quad SR\_STATE(i,l+1) \leftarrow A$
and $SR\_CROSS(i,l+1) \leftarrow A$
$\forall i$ where $SR\_CROSS(i,l) = B$ and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) = 1$, and $I\_SR\_LOW(i)$:
$SR\_CROSS(i,l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i,l) \neq B$, $SR\_FNUM(i,l) \geq 2$, and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) \neq B$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$T \leftarrow T-1$

ELSE
$\forall i \in B\_ILLEFT(T-1,T): \quad IL\_STATE(i,l+1) \leftarrow A$

IF $B\_FCON(T-1,T) \neq \emptyset$ **THEN**
Follow instructions given by $RULES\_SRDOWNSR\_FREQ$, except $T \leftarrow T-2$.

ELSE
Follow $RULES\_SRDOWNSR\_NOFREQ$, and the exception, $T \leftarrow T-2$.
END
END
END

IF $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_HIGH \quad \wedge \quad R\_TYPE(T+1) = SR$
$\wedge \quad B\_FCON(T,T+1) \neq \emptyset$

($RULES\_SRUPSR\_FREQ$)
THEN $\forall i \in B\_SRFINISH(T,T+1): \quad SR\_STATE(i,l+1) \leftarrow B$ and $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i \in B\_SRSTART(T,T+1): \quad SR\_STATE(i,l+1) \leftarrow C$
$\forall i \in B\_SRSTART(T,T+1)$ and $u_{i,l} = 0$: $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i \in B\_SRSTART(T,T+1)$ and $u_{i,l} \neq 0$: $SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$

$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i$ where $SR\_CROSS(i,l) = S$, $i \notin B\_SRFINISH(T,T+1)$, and $SR\_FNUM(i,l) \geq 2$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = B$, $i \notin B\_SRFINISH(T,T+1)$, and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) = 2$, $i \notin B\_SRFINISH(T,T+1)$,
and $I\_SR\_LOW(i)$: $SR\_CROSS(i,l+1) \leftarrow B$ and $SR\_FNUM(i,l+1) \leftarrow 1$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) \geq 3$, and $i \notin B\_SRFINISH(T,T+1)$,
and $I\_SR\_LOW(i)$: $SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) \geq 2$, and $i \notin B\_SRFINISH(T,T+1)$,
and $I\_SR\_HIGH(i)$: $SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$T \leftarrow T+1$

END

IF $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_HIGH \quad \wedge \quad R\_TYPE(T+1) = SR$
$\wedge \quad B\_FCON(T,T+1) = \emptyset$

$(RULES\_SRUPSR\_NOFREQ)$
THEN $\forall i \in B\_SRFINISH(T,T+1)$: $SR\_STATE(i,l+1) \leftarrow B$
and $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i \in B\_SRSTART(T,T+1)$: $SR\_STATE(i,l+1) \leftarrow C$
$\forall i \in B\_SRSTART(T,T+1)$ and $u_{i,l} = 0$: $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i \in B\_SRSTART(T,T+1)$ and $u_{i,l} \neq 0$: $SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i$ where $SR\_CROSS(i,l) = B$, $i \notin B\_SRFINISH(T,T+1)$, and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = S$, $i \notin B\_SRFINISH(T,T+1)$, $SR\_FNUM(i,l) = 1$,
and $I\_SR\_LOW(i)$: $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i,l) = S$, $i \notin B\_SRFINISH(T,T+1)$, $SR\_FNUM(i,l) \geq 2$,
and $I\_SR\_LOW(i)$: $SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = S$, $i \notin B\_SRFINISH(T,T+1)$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$T \leftarrow T+1$

END

IF $R\_TYPE(T) = SR \quad \wedge \quad I\_TIME\_HIGH \quad \wedge \quad R\_TYPE(T+1) = IL$

THEN $\forall i \in B\_ILRIGHT(T, T+1): \quad IL\_STATE(i, l+1) \leftarrow C$

Find the total IL dispatched at this boundary:

$$IL\_TOT = \sum_{\substack{i \\ IL\_STATE(i,l+1)=C}} p_{i,l}$$

IF $IL\_TOT > 0 \quad \vee \quad \nexists i$ where $SR\_STATE(i, l) \neq B$ and $i \notin B\_SRFINISH(T, T+1)$

THEN $\forall i \in B\_SRFINISH(T, T+1): \quad SR\_STATE(i, l+1) \leftarrow B$
and $SR\_CROSS(i, l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i) = A$ and $I\_SR\_LOW(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i) = A$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$

IF $B\_FCON(T, T+1) \neq \emptyset$
THEN $\forall i$ where $SR\_CROSS(i, l) = B$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_HIGH(i): \quad SR\_CROSS(i, l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i) = S$, $SR\_FNUM(i) \geq 2$, and $i \notin B\_SRFINISH(T, T+1)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i) = S$, $SR\_FNUM(i) = 2$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_LOW(i): \quad SR\_CROSS(i) \leftarrow B$
$\forall i$ where $SR\_CROSS(i) = S$, $SR\_FNUM(i) \geq 3$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_LOW(i): \quad SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i) = S$, $SR\_FNUM(i) \geq 2$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_HIGH(i): \quad SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$T \leftarrow T + 1$

ELSE
$\forall i$ where $SR\_CROSS(i, l) = B$, $i \notin B\_SRFINISH(T, T+1)$, and $I\_SR\_HIGH(i)$:
$SR\_CROSS(i, l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i, l) = S$, $SR\_FNUM(i, l) = 1$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_LOW(i): \quad SR\_CROSS(i, l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i, l) = S$, $SR\_FNUM(i, l) \geq 2$, $i \notin B\_SRFINISH(T, T+1)$,
and $I\_SR\_LOW(i): \quad SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$
$\forall i$ where $SR\_CROSS(i, l) = S$, $i \notin B\_SRFINISH(T, T+1)$, and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$
$T \leftarrow T + 1$

END

ELSE

$\forall i \in B\_ILRIGHT(T, T+1): \quad IL\_STATE(i, l+1) \leftarrow C$

IF $B\_FCON(T, T+1) \neq \emptyset$ **THEN**

Follow $RULES\_SRUPSR\_FREQ$, and the exception $T \leftarrow T+2$.

ELSE

Follow $RULES\_SRUPSR\_NOFREQ$, and the exception $T \leftarrow T+2$.

END

END

END

IF $\neg I\_TIME\_LOW \quad \wedge \quad \neg I\_TIME\_HIGH$

THEN $\forall i$ where $SR\_CROSS(i, l) = B$ and $I\_SR\_HIGH(i)$:

$SR\_CROSS(i, l+1) \leftarrow S$

$\forall i$ where $SR\_CROSS(i, l) = S$, $SR\_FNUM(i, l) = 1$, and $I\_SR\_LOW(i)$:

$SR\_CROSS(i, l+1) \leftarrow B$

$\forall i$ where $SR\_CROSS(i, l) = S$, $SR\_FNUM(i, l) \geq 2$, and $I\_SR\_LOW(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$

$\forall i$ where $SR\_CROSS(i, l) = S$ and $I\_SR\_HIGH(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$

END

IF $R\_TYPE(T) = IL \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad \neg I\_TIME\_HIGH$
$\wedge \quad B\_FCON(T-1, T) \neq \emptyset$

THEN $\forall i \in B\_ILRIGHT(T-1, T): \quad IL\_STATE(i) \leftarrow A$

$\forall i$ where $SR\_CROSS(i, l) = A$ and $I\_SR\_LOW(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$

$\forall i$ where $SR\_CROSS(i, l) = A$ and $I\_SR\_HIGH(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$

$\forall i$ where $SR\_CROSS(i, l) = B$ and $I\_SR\_HIGH(i)$:

$SR\_CROSS(i, l+1) \leftarrow S$ and $SR\_FNUM(i, l+1) \leftarrow 2$

$\forall i$ where $SR\_CROSS(i, l) = S: \quad SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$

$\forall i$ where $SR\_CROSS(i, l) = S$ and $I\_SR\_LOW(i)$:

$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) - 1$

$\forall i$ where $SR\_CROSS(i,l) = S$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i \in B\_SRFINISH(T-1,T)$: $SR\_STATE(i,l+1) \leftarrow C$
$\forall i$ where $i \in B\_SRFINISH(T-1,T)$ and $u_{i,l} = u_i^{max}$: $SR\_CROSS(i) \leftarrow S$
$T \leftarrow T-1$

END

IF $R\_TYPE(T) = IL \quad \wedge \quad I\_TIME\_LOW \quad \wedge \quad \neg I\_TIME\_HIGH$
$\wedge \quad B\_FCON(T-1,T) = \emptyset$

THEN $\forall i \in B\_ILRIGHT(T-1,T)$: $IL\_STATE(i) \leftarrow A$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i$ where $SR\_CROSS(i,l) = B$ and $I\_SR\_HIGH(i)$: $SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = S$ and $I\_SR\_LOW(i)$: $SR\_CROSS(i,l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) \geq 2$, and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = S$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i \in B\_SRFINISH(T-1,T)$: $SR\_STATE(i,l+1) \leftarrow C$
$\forall i$ where $i \in B\_SRFINISH(T-1,T)$ and $u_{i,l} = u_i^{max}$: $SR\_CROSS(i) \leftarrow S$
$T \leftarrow T-1$

END

IF $R\_TYPE(T) = IL \quad \wedge \quad I\_TIME\_HIGH$

THEN $\forall i \in B\_ILLEFT(T,T+1)$: $IL\_STATE(i) \leftarrow B$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = A$ and $I\_SR\_HIGH(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) + 1$
$\forall i$ where $SR\_CROSS(i,l) = B$ and $I\_SR\_HIGH(i)$: $SR\_CROSS(i,l+1) \leftarrow S$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) = 1$, and $I\_SR\_LOW(i)$:
$SR\_CROSS(i,l+1) \leftarrow B$
$\forall i$ where $SR\_CROSS(i,l) = S$, $SR\_FNUM(i,l) \geq 2$, and $I\_SR\_LOW(i)$:
$SR\_FNUM(i,l+1) \leftarrow SR\_FNUM(i,l+1) - 1$
$\forall i$ where $SR\_CROSS(i,l) = S$ and $I\_SR\_HIGH(i)$:

$$SR\_FNUM(i, l+1) \leftarrow SR\_FNUM(i, l+1) + 1$$
$$\forall i \in B\_SRSTART(T, T+1): \quad SR\_STATE(i, l+1) \leftarrow C$$
$$\forall i \text{ where } i \in B\_SRSTART(T, T+1) \text{ and } u_{i,l} > 0: \quad SR\_CROSS(i, l+1) \leftarrow S$$
$$\forall i \text{ where } i \in B\_SRSTART(T, T+1) \text{ and } u_{i,l} = 0: \quad SR\_CROSS(i, l+1) \leftarrow B$$
$$T \leftarrow T + 1$$

END

B.4 GLOBAL MINIMUM CHECKING

This section describes the algorithm used to determine if a point is the global minimum or not, Line 8 of Main Algorithm. To appreciate its purpose, a simplified example is explained. Consider a general convex solution space shown in Figure B.1 with different cost vectors. The example solution space is divided into regions, by two internal boundaries, creating four regions. Three possible cost vectors, $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, are positioned at the global minimum for that vector.

Consider the problem of Figure B.1 with cost vector $\mathbf{c}_1$. Starting in region (b), the local minimum is found at the intersection of the two internal boundaries. The transition rules of Section B.3 specify that (c) should be solved next, not (a) or (d), because (c) is directly opposite. The local minimum of (c) is found, indicated by a dot in Figure B.1 by the $\mathbf{c}_1$ vector. The minimum does not lie on any internal boundary, and therefore has to be the global minimum. This is the standard process of solving the problem, applying the same process to $\mathbf{c}_2$ a similar result is obtained, a minimum is found on the vertex, and the algorithm terminates. However, for $\mathbf{c}_3$ a difficultly arises: starting in region (b), the local minimum is found on the top left vertex. Since this point lies on an internal boundary, region (a) is checked and the local minimum is found on the top vertex of it. If the same transition rules are applied then region (b) is optimised again, and the process will continue ad infinitum.

Although situations like $\mathbf{c}_3$ do not occur often they are a possibility and need to be accounted for. In a two dimensional problem whenever this situation arises the point of interest is always the global minimum, but in higher dimensions this is not always the case, and it has to be decided whether it is global or not, this is the goal of this section. The next section determines what to do when it is not the global minimum.

For a linear objective function minimising over a bounded convex space, the minimum has to lie on the external boundary. Therefore there has to be at least one external boundary, as defined in Section 5.3.4 and generalised here by $h_j(\mathbf{v}) \leq 0$, where $h_j(\mathbf{v}^*) = 0$. In most situations there is more than one such external boundary, but due to the non-linear nature of this optimisation it is possible just to have one. After further reviewing the first solved example, it can be shown as an instance of this, the example from Section 5.4.1. Since there are many points on external boundaries, $h_j(\mathbf{v}^*) = 0$ is

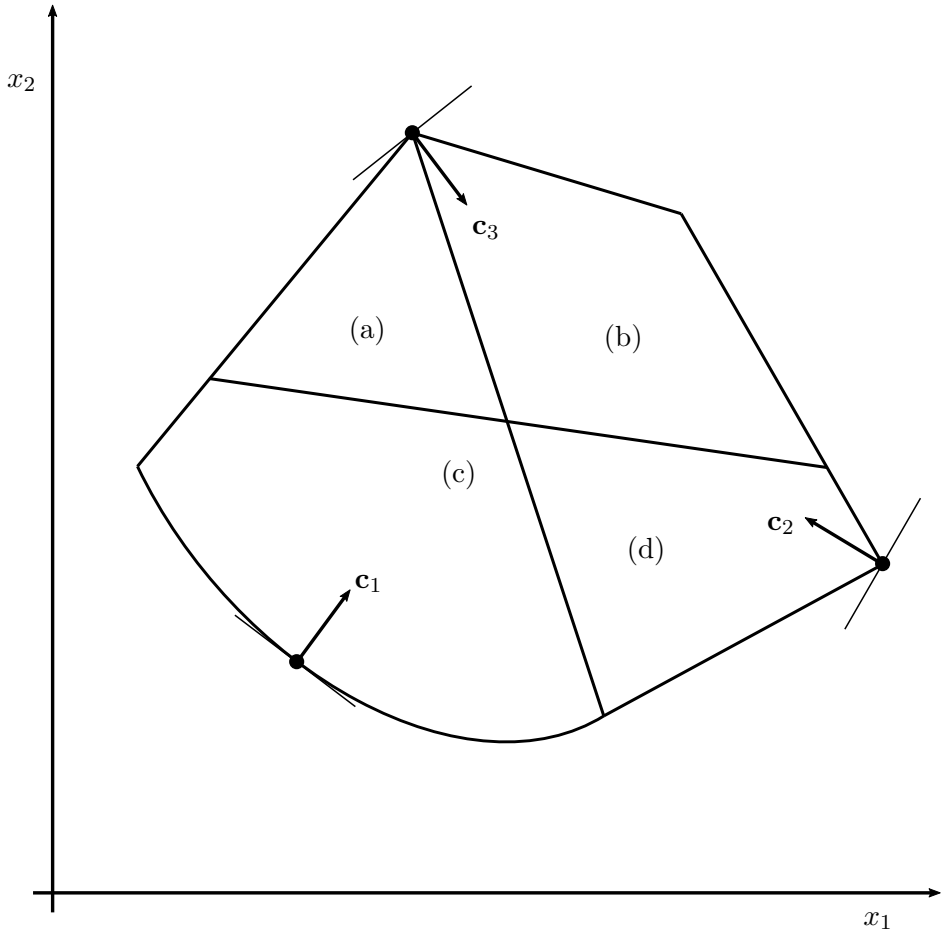


Figure B.1 Example of possible places where the global minimum can be found.

not enough to know the global minimum has been found. Therefore a second necessary condition is required and is obtained from the KKT conditions:

$$\mathbf{c} + \sum_{\substack{j \\ h_j(\mathbf{v}^*)=0}} \mu_j \nabla h_j(\mathbf{v}^*) = 0 \quad \text{and} \quad \mu_j \geq 0 \tag{B.1}$$

where $\mathbf{c}$ is the general cost vector, ∇ is the del operator working on the external constraint function, with the result of a vector evaluated at the minimum. It should be mentioned about the components of $\mathbf{c}$ and $\mathbf{v}$, that the main variables, p_i and u_i, are the only components. If the optimisation is extended to included risk and inertia, then R and H are added in. However t_{min}, p_m, and any slack variables are not included, although they may appear for the purposes of simplification in the description of $h_j(\mathbf{v})$.

If there exists μ_j such that Eq. B.1 is satisfied, then it is known that $\mathbf{v}^*$ is the global minimum. This equation is a linear problem, μ_j are the variables and identified as KKT multipliers. The solving methodology will not be explained in detail, but it

is more complicated than direct matrix inversion, rather a row reduction approach is recommended, with additional handling for infeasible and non-unique problems.

B.4.1 Gradient Clarification

A clarification is required for which external constraints, $h_j(\mathbf{v})$, are applied in Eq. B.1. Although $h_j(\mathbf{v}^*) = 0$ is a necessary condition, it is not a complete description in relation to the minimum frequency constraint, which suffers from discontinuities. For reserve limits, Eq. 5.13 and 5.14, and reserve requirement constraint, Eq. 5.7, $h_j(\mathbf{v})$ is independent of region, and $\nabla h_j(\mathbf{v}^*)$ can be found without any difficultly. The same cannot be implied about the frequency limits and minimum frequency constraints, therefore the definition of $h_j(\mathbf{v})$ for these constraints is the first step.

1. Frequency Limits

   Define $h_j(\mathbf{v})$ for a frequency limit in the domain of each region by rearranging Eq. 5.83 and 5.90, for IL and SR type regions respectively:

$$h_j(\mathbf{v}) = 2Hf_j + Rt_{j,l} - \sum_{i \in Q_{B,j}} p_i(t_{j,l} - t_{i,p}) - \sum_{i \in W_{B,j}} \left(u_i(t_{j,l} - t_{i,u}) - \frac{u_i^2}{2g_i}\right) - \sum_{i \in W_{T,j}} \frac{g_i}{2}(t_{j,l} - t_{i,u})^2 \tag{B.2}$$

   It is noticed that the equation's form only changes when a SR offer transitions from $W_{B,j}$ to $W_{T,j}$ or vice versa. It does this when $u_i = g_i(t_{j,l} - t_{i,u})$, and when this occurs:

$$u_i(t_{j,l} - t_{i,u}) - \frac{u_i^2}{2g_i} = \frac{g_i}{2}(t_{j,l} - t_{i,u})^2$$

   Therefore $h_j(\mathbf{v})$ is continuous. The components of $\nabla h_j(\mathbf{v})$ vector are:

$$\nabla h_{i,j}(\mathbf{v}) = \begin{cases} -(t_{j,l} - t_{i,p}) & \text{if IL offer and} \quad i \in Q_{B,j} \\ 0 & \text{if IL offer and} \quad i \notin Q_{B,j} \\ -(t_{j,l} - t_{i,u}) + \frac{u_i}{g_i} & \text{if SR offer and} \quad i \in W_{B,j} \\ 0 & \text{if SR offer and} \quad i \notin W_{B,j} \end{cases} \tag{B.3}$$

   Therefore $\nabla h_j(\mathbf{v})$ is also continuous in each component, including SR offers when on the boundary between being $W_{B,j}$ and $W_{T,j}$, $u_i = g_i(t_{j,l} - t_{i,u})$, $\nabla h_{i,j}(\mathbf{v}) = 0$. Hence applying frequency limits to Eq. B.1 is entirely dependent on whether $h_j(\mathbf{v}^*) = 0$.

2. Minimum Frequency Constraint

Defined $h_j(\mathbf{v})$ for the minimum frequency constraint in each region by substituting Eq. 5.81 into Eq. 5.82 for IL type region, and Eq. 5.88 into Eq. 5.89 for SR type region, and then rearranging to obtain:

$$h_j(\mathbf{v}) = 2Hf_{lim}+Rt_{min}-\sum_{i\in Q_B} p_i(t_{min}-t_{i,p})-\sum_{i\in W_B}\Big(u_i(t_{min}-t_{i,u})-\frac{u_i^2}{2g_i}\Big)-\sum_{i\in W_T}\frac{g_i}{2}(t_{min}-t_{i,u})^2 \tag{B.4}$$

Although this equation is similar in style to Eq. B.2, its form is significantly different due to $t_{min} = t_{min}(\mathbf{v})$ and $f_{lim} = f_{lim}(t_{min}(\mathbf{v}))$. It has been shown in Appendix A for t_{min}, and for f_{lim} in Eq. 5.5, that they can be discontinuous, and by extension $h_j(\mathbf{v})$ is also. Discontinuity is only found across internal boundaries where index T changes number. A change in region where only SR offers interchange between W_B and W_T does not cause a discontinuity on a internal boundary.

$\nabla h_j(\mathbf{v})$ is found in each region:

$$\nabla h_{i,j}(\mathbf{v}) = \begin{cases} -(t_{min}-t_{i,p}) & \text{if IL offer and} \quad i\in Q_B \\ 0 & \text{if IL offer and} \quad i\notin Q_B \\ -(t_{min}-t_{i,u})+\frac{u_i}{g_i} & \text{if SR offer and} \quad i\in W_B \\ 0 & \text{if SR offer and} \quad i\notin W_B \end{cases} \tag{B.5}$$

This formula is not derived by assuming t_{min} or p_m are constant, as it may appear, but is derived from the equations that form Eq. B.4 and applying the chain rule. $\nabla h_j(\mathbf{v})$ depends on t_{min} so each of its components is not necessarily continuous as well. The consequence this has on formulating Eq. B.1 depends on whether at $\mathbf{v}^*$, $h_j(\mathbf{v}^*)$ and $t_{min}(\mathbf{v}^*)$ are continuous in all directions. If so then the procedure follows the standard means, as it does for the other constraints. If not, then execute the following:

- For each region where $\mathbf{v}^*$ is on its border, find all t_{min} defined by each region at $\mathbf{v}^*$. Each unique instance of t_{min} can be ordered from earliest to latest. Order by the index, m, to get the series: $t_{min,1}$, $t_{min,2}$, $\cdots$, $t_{min,Nm}$.
- If $t_{min,m}$ is not equal to the time of any frequency limits, $t_{j,l}$, and $h_j(\mathbf{v}^*) = 0$ for the specific region associated with $t_{min,m}$, then add the vector $\nabla h_{j,m}(\mathbf{v}^*)$ to Eq. B.1.

For the minimum frequency constraint, due to its discontinuity, more than one vector can be added to Eq. B.1. The reason why the gradient vector, $\nabla h_{j,m}(\mathbf{v}^*)$, is omitted when there exists j where $t_{min,m} = t_{j,l}$ is because there will be

a corresponding frequency limit vector that will be exactly the same. It is unnecessary to have multiple instances of that same vector.

So far this description has been quite abstract, therefore an example is shown when a discontinuity in the minimum frequency constraint occurs. A set of offers have been produced, a set of frequency limits have been created, and the optimisation process has been executed for a number of steps. Currently two consecutive regions have resulted in the exact same local minimum $\mathbf{v}^*$. The frequency transient for this local minimum is shown in Figure B.2.

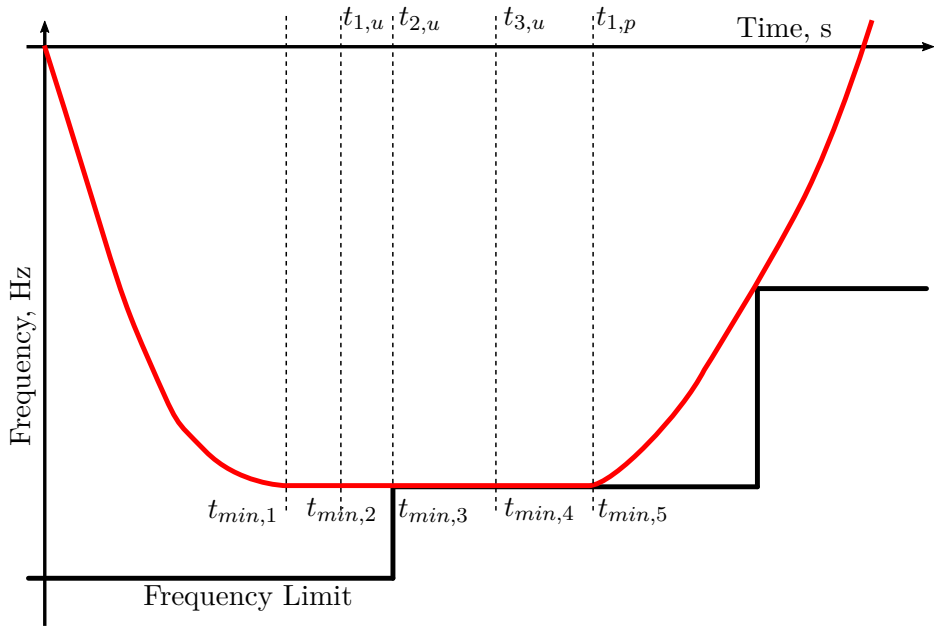


Figure B.2 Example of a frequency transient where the minimum frequency constraint function is discontinuous.

When multiple instances of $t_{min,m}$ occur it is because the frequency transient is flat at the minimum. According to the definition of t_{min} in Eq. 5.9 and 5.10, $t_{min,1}$ is the proper minimum time, but for this $\mathbf{v}^*$, t_{min} can jump instantaneously to any of the others. Other times are from when more reserve can be dispatched, but emphasis is put on 'can' as u_1^*, u_2^*, and u_3^* are all dispatched to zero for this local minimum, because the frequency transient is flat. Two vectors, $\nabla h_{j,m}(\mathbf{v}^*)$, are only added for the fourth and fifth $t_{min,m}$. For the first and second, the frequency transient does not come up against the limit, i.e. $h_j(\mathbf{v}^*) \neq 0$, and for the third, it is the time of a frequency limit step change. Therefore two vectors are added in Eq. B.1 for the minimum frequency constraint, a distinction that does not occur for any of the other constraints.

B.4.2 Nonuniqueness Clarification

The conditional statement of Line 7 of Main Algorithm is split into two parts, the first when current and previous solutions were both unique, and the second when one of them at least is not unique with an addition expression. The expression $\{\mathbf{v}_l\} \cap \{\mathbf{v}_{l-1}\} \neq \emptyset$ means when both solutions share some points. However, this is difficult to compute in a practical algorithm, hence it is avoided. Instead the second part is omitted from the actual implementation, and on Line 11 of Main Algorithm a check is added. This check determines if the next region has been optimised before, and if it has it jumps to line 10, and makes a second choice for the next region.

B.5 SECOND CHOICE TRANSITION RULES

The second set of transition rules is interested in the geometric structure and direction of least cost, not making assumptions on the optimal region to pick next, but calculating. This method determines the direction of least cost along the external boundary upon which the local minimum lies. Then the region which this direction points to is chosen for optimisation. This process is computationally more difficult than the first set of transition rules and so is only applied for difficult situations.

The cost vector, $\mathbf{c} = [\mathbf{c}_p^T, \mathbf{c}_u^T]^T$, points in the direction of maximum increasing cost, therefore $-\mathbf{c}$ points to where the next region should be. However, if the common local minimum, $\mathbf{v}^*$, lies on external boundary where $-\mathbf{c}$ points outside the feasible space, then a direction along the boundary is required. Define $\mathbf{e}$ as the heading of least cost. In the absence of the local minimum lying on any external boundaries, $\mathbf{e} = -\mathbf{c}$. For local minimums that do lie on external boundaries, then $\mathbf{e}$ is the projection of $-\mathbf{c}$ onto the plane of those boundaries. For the local minimum, $\mathbf{v}^*$, there maybe several intersected boundaries, where $h_j(\mathbf{v}^*) = 0$. A movement from $\mathbf{v}^*$ will most likely result in one, $h_j(\mathbf{v})$, being equal to zero, although this is not necessary.

The principle of projecting $-\mathbf{c}$ onto the external boundary, is to minimise the angle between $\mathbf{e}$ and $-\mathbf{c}$, the equation for angle θ is:

$$-\mathbf{c} \bullet \mathbf{e} = |\mathbf{c}||\mathbf{e}| \cos(\theta) \tag{B.6}$$

It is inconvenient to minimise inverse cosine formula, instead an equivalent optimisation is made:

$$\text{maximise} \quad \frac{-\mathbf{c} \bullet \mathbf{e}}{|\mathbf{c}||\mathbf{e}|} \tag{B.7}$$

The projection $\mathbf{e}$ is restricted to be within the external boundaries, this constraint

is expressed as:

$$\forall j \quad \text{where} \quad h_j(\mathbf{v}^*) = 0 \quad \text{then} \quad \nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} \leq 0 \tag{B.8}$$

The choice of $h_j(\mathbf{v}^*)$ and $\nabla h_j(\mathbf{v}^*)$ requires the same considerations as Section B.4.1. The current problem is not well formulated yet, as the solution is not unique: let $\mathbf{e}_0$ be a solution to this problem, then for any $\alpha > 0$, $\alpha\mathbf{e}_0$ is also an optimal solution. Therefore an extra condition is added to normalise $\mathbf{e}$, $|\mathbf{e}|^2 = 1$. The optimisation problem becomes:

$$\text{minimise} \quad \mathbf{c} \bullet \mathbf{e} \tag{B.9}$$

$$\text{subject to} \quad \mathbf{e} \bullet \mathbf{e} = 1 \tag{B.10}$$

$$\text{and} \quad \forall j \in E \quad \nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} \leq 0 \tag{B.11}$$

where E is the set of critical external boundaries determined by Eq. B.8 and Section B.4.1. To find the optimal solution the KKT conditions can be solved:

$$\mathbf{c} + 2\lambda\mathbf{e} + A^T\boldsymbol{\mu} = 0 \tag{B.12}$$

$$\mathbf{e} \bullet \mathbf{e} - 1 = 0 \tag{B.13}$$

$$A\mathbf{e} \otimes \boldsymbol{\mu} = 0 \tag{B.14}$$

$$A\mathbf{e} \leq 0 \tag{B.15}$$

$$\boldsymbol{\mu} \geq 0 \tag{B.16}$$

where λ and $\boldsymbol{\mu}$ are the KKT multipliers for Eq. B.10 and B.11 respectively. The operator, $\otimes$, is the element-wise multiplication. The matrix A is:

$$A = \begin{bmatrix} \nabla h_1(\mathbf{v}^*)^T \\ \nabla h_2(\mathbf{v}^*)^T \\ \vdots \\ \nabla h_j(\mathbf{v}^*)^T \end{bmatrix} \tag{B.17}$$

Details about solving these equations will not be given, except that a Newton's method would be possible. One simplification is appropriate, the removal of λ from Eq. B.12. Firstly, dot multiply Eq. B.12 with $\mathbf{e}$:

$$\mathbf{c} \bullet \mathbf{e} + 2\lambda\mathbf{e} \bullet \mathbf{e} + (A^T\boldsymbol{\mu}) \bullet \mathbf{e} = 0$$

It is known from Eq. B.14 that:

$$(A^T\boldsymbol{\mu}) \bullet \mathbf{e} = (A\mathbf{e}) \bullet \boldsymbol{\mu} = (A\mathbf{e} \otimes \boldsymbol{\mu}) \bullet \mathbf{1} = 0$$

Therefore $\lambda = -(\mathbf{c} \bullet \mathbf{e})/2$ and Eq. B.12 is transformed to:

$$\mathbf{c} - (\mathbf{c} \bullet \mathbf{e})\mathbf{e} + A^T\boldsymbol{\mu} = 0 \tag{B.18}$$

The next step is to determine which region $\mathbf{e}$ points to. For the shared local minimum $\mathbf{v}^*$ there is a series of regions that border on that point, these regions are indexed by r. For each region there is a set of internal boundaries, $\mathbf{a}_{i,r} \bullet \mathbf{v} \leq b_{i,r}$, where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$ for at least one i. Choose the region, r, where for all i when $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$, then $\mathbf{a}_{i,r} \bullet \mathbf{e} \leq 0$.

A special constraint is required when $\mathbf{a}_{i,r} \bullet \mathbf{e} = 0$, the requirement is that $\mathbf{d} \bullet \mathbf{a}_{i,r} \leq 0$, where $\mathbf{d}$ points inwards into the feasible space from the external boundaries at $\mathbf{v}^*$, and is not equal to $\mathbf{e}$. More precisely, for each external boundary where $h_j(\mathbf{v}^*) = 0$ and $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$, then $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{d} < 0$. Notice it is not required for all external constraints where $h_j(\mathbf{v}^*) = 0$ that $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{d} < 0$, as it is not necessary for ones where $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} < 0$. The direction, $\mathbf{d}$, is interpreted as pointing inwards into the feasible solution space along the local direction of $\mathbf{e}$. The necessity of this constraint has not been demonstrated from the above discussion. However in the next section, the reason will become evident while demonstrating its existence.

The next step is to construct $\mathbf{d}$, according the requirements above. These conditions do not uniquely define $\mathbf{d}$, so different methods are possible, which can create different choices for $\mathbf{d}$. An approach is to construct $\mathbf{d}$ as a linear combination of critical external boundaries:

$$\mathbf{d} = \sum_{\substack{j \\ \nabla h_f(\mathbf{v}^*) \bullet \mathbf{e} = 0}} \beta_j \nabla h_f(\mathbf{v}^*) \tag{B.19}$$

where β_j are parameters adjusted to satisfy conditions. To arrive at one solution, set $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{d} = -1$ in order to satisfy $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{d} < 0$, to form a series of linear equations:

$$\begin{bmatrix} \nabla h_1 \bullet \nabla h_1 & \nabla h_1 \bullet \nabla h_2 & \cdots \\ \nabla h_2 \bullet \nabla h_1 & \nabla h_2 \bullet \nabla h_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \vdots \end{bmatrix} \tag{B.20}$$

The solution for β_j becomes a matter of solving this set of linear equations. In some rare instances, the matrix may be ill-conditioned and require further equations and restrictions on the right hand side.

The set of rules in this section are also applied for non-unique local minimums. Some details will change in some instances but the same principles apply. This marks

the completion of the region transition algorithm.

B.6 TERMINATION

For a practical solver, it has to be known whether it will stop. Therefore this section will demonstrate that the algorithm will terminate. Firstly, it is known for any region the local minimum can be found, but it has not been shown whether the region transition algorithm will stop yet. In Section 5.3.1, it has been demonstrated that there are a finite number of regions, therefore solving each region will eventually arrive at the correct answer, hence there exists known methods that will terminate. This method is prohibitively slow given the total number of regions, Eq. 5.98, so a faster algorithm has been developed.

The only possible means for this algorithm not to terminate is if it oscillates between one set of regions, which all share the same local minimum. The Main Algorithm in Section B.2 is designed to avoid this situation. Line 7 of Main Algorithm checks whether the local minimum has been repeated and straight away checks if this point is the global minimum. If it is not, then it makes sure the next region will have a smaller local minimum, Line 10 of Main Algorithm. This avoids the repetition of regions and guarantees the algorithm will stop. The goal of this section is to demonstrate that Section B.5 will produce a region with a smaller local minimum than the previous one.

To prove that the new region, indexed by r, has a local minimum smaller than the previous solution, it has to be shown that the intersection of the following three sets is not empty, $C_r = S_F \cap Z_r \cap C(\mathbf{v}^*) \neq \emptyset$, where:

- S_F, the set of all feasible points determined by the external boundaries.
- Z_r, the set of all points that satisfy the internal boundaries of region r. Its proper definition is made in Eq. 5.127.
- $C(\mathbf{v}^*)$, the set of all points with a cost less than the current local minimum: $C(\mathbf{v}^*) = \{\mathbf{v} \in \mathbb{R}^N : \mathbf{c} \bullet (\mathbf{v} - \mathbf{v}^*) < 0\}$.

The first two sets define the region, whereas the third specifies all the points with a value less than the current local minimum. If this set is not empty, then the new region will definitely have a local minimum less than the previous one, and prove the algorithm will terminate. This proof is made by a method of construction, that is to show it is always possible to construct a point that is a member of that set C_r.

Firstly, for the simple case when $\mathbf{e} = -\mathbf{c}$ with no external boundaries where $h_j(\mathbf{v}^*) = 0$ and $\mathbf{c} \bullet \nabla h_j(\mathbf{v}^*) = 0$, then choose an $\alpha > 0$ where $\mathbf{v}^* + \alpha\mathbf{e}$ satisfies all internal and external boundaries:

- For internal boundaries where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$, then $\mathbf{a}_{i,r} \bullet (\mathbf{v}^* + \alpha\mathbf{e}) = \mathbf{a}_{i,r} \bullet \mathbf{v}^* + \alpha(\mathbf{a}_{i,r} \bullet \mathbf{e}) \leq b_{i,r}$ because $\mathbf{a}_{i,r} \bullet \mathbf{e} \leq 0$. Therefore any $\alpha > 0$ is sufficient.

- For other internal boundaries, it is required that $\mathbf{a}_{i,r} \bullet \mathbf{v}^* \leq b_{i,r}$, otherwise $\mathbf{v}^*$ is not an element of that region. Then α has to be less than:

$$\alpha \leq \frac{b_{i,r} - \mathbf{a}_{i,r} \bullet \mathbf{v}^*}{\mathbf{a}_{i,r} \bullet \mathbf{e}} \tag{B.21}$$

  when $\mathbf{a}_{i,r} \bullet \mathbf{e}$ is positive, if negative then α would have to be greater than a negative number which means any $\alpha > 0$ is sufficient.

- For external boundaries, there are two possible situations: (a) $h_j(\mathbf{v}^*) < 0$ or (b) $h_j(\mathbf{v}^*) = 0$ and $\mathbf{c} \bullet \nabla h_j(\mathbf{v}^*) > 0$. To determine what happens for each situation consider the Taylor Series expansion:

$$h_j(\alpha) = h_j(\mathbf{v}^*) + \alpha(\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e}) + \mathcal{O}(\alpha^2) \tag{B.22}$$

  For situation (a) an approximate maximum for α can be found, provided $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} > 0$:

$$\alpha \lesssim -\frac{h_j(\mathbf{v}^*)}{\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e}} \tag{B.23}$$

  The exact limit will depend on the quadratic term, but it is known that there will be a positive limit as α can be made smaller to mitigate the effect of the quadratic term. If $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} \leq 0$ then α should be kept small enough so that $h_j(\alpha) < 0$, i.e. before the quadratic term dominates over the $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e}$ term; the same applies to situation (b).

Secondly, for the complex cases, where there is at least one external boundary constraint such that $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$, the process of finding a point in C_r is more difficult. It is not guaranteed that any point along $\mathbf{v}^* + \alpha\mathbf{e}$ for $\alpha > 0$ will be possible member of C_r, especially if h_j is for one of the frequency constraints, as the coefficient of the quadratic term, $\mathcal{O}(\alpha^2)$ in Eq. B.22, is positive for the feasible space to be convex. Therefore $h_j(\alpha) > 0$ and any value of α will rest on a point outside the feasible region. Another approach is required.

A new direction is found by making a small adjustment to $\mathbf{e}$, but before this, it is convenient to scale $\mathbf{e}$ by $-\mathbf{c} \bullet \mathbf{e}$, so that $\mathbf{f} = -(\mathbf{c} \bullet \mathbf{e})\mathbf{e}$. Therefore rearranging Eq. B.18, a general formula for $\mathbf{f}$ is given:

$$\mathbf{f} = -\mathbf{c} - A^T\boldsymbol{\mu} \tag{B.24}$$

A couple of remarks should be made. Firstly, $-\mathbf{c} \bullet \mathbf{e}$ is positive and does not change the direction of $\mathbf{e}$. If the opposite were true, $\mathbf{c} \bullet \mathbf{e}$ is positive, then this would imply that the only possible feasible points around $\mathbf{v}^*$ would result in $\mathbf{c} \bullet (\mathbf{v}^* + \alpha\mathbf{e}) > \mathbf{c} \bullet \mathbf{v}^*$ and imply $\mathbf{v}^*$ was the global minimum, which it is not. If it was the global minimum then there would not be any need for these rules, as the process would have terminated already.

Secondly, the nature of $\boldsymbol{\mu}$ from Eq. B.14 is that either $\mu_j = 0$ or $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$. The significance of this constraint is that there is a set of external boundaries where $\mathbf{e}$ lies along, and a set it does not. Call the first set J_0 and the second one J_1. They correspond to $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$ and $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} < 0$ respectively. It is more helpful to view Eq. B.24 in terms of J_0:

$$\mathbf{f} = -\mathbf{c} - \sum_{j \in J_0} \mu_j \nabla h_j(\mathbf{v}^*) \tag{B.25}$$

This provides an interpretation to $\mathbf{f}$ as the backwards projection of $-\mathbf{c}$ via the normals of each critical external boundary, $\nabla h_j(\mathbf{v}^*)$. To change the direction of $\mathbf{f}$ and therefore $\mathbf{e}$, a good approach would be to extend the backwards projection, but modify it so that it remains within the external boundaries:

$$\mathbf{g} = \mathbf{f} + \eta\mathbf{d} \tag{B.26}$$

where $\mathbf{d}$ is defined in Eq. B.19 and points inwards into the feasible space, and η is parameter determining how modified $\mathbf{f}$ is is by $\mathbf{d}$. The next step is to determine the limit on η so that $\mathbf{g}$ still points within C_r. Firstly consider $\mathbf{c} \bullet \mathbf{g} < 0$ so that $\mathbf{v}^* + \alpha\mathbf{g}$ will be an element of $C(\mathbf{v}^*)$. Substitute $\mathbf{f} = -(\mathbf{c} \bullet \mathbf{e})\mathbf{e}$ into Eq. B.26 and dot product with $\mathbf{c}$:

$$\mathbf{c} \bullet \mathbf{g} = -(\mathbf{c} \bullet \mathbf{e})^2 + \eta\mathbf{d} \bullet \mathbf{c} < 0$$

There is a limit on η if $\mathbf{d} \bullet \mathbf{c} > 0$:

$$\eta < \frac{(\mathbf{c} \bullet \mathbf{e})^2}{\mathbf{d} \bullet \mathbf{c}} \tag{B.27}$$

if $\mathbf{d} \bullet \mathbf{c} \leq 0$, then any value of $\eta > 0$ will satisfy $\mathbf{c} \bullet \mathbf{g} < 0$.

The next set of limits on η is from internal boundaries where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$. The requirement is $\mathbf{a}_{i,r} \bullet \mathbf{g} \leq 0$ so that for $\alpha > 0$ then $\mathbf{a}_{i,r} \bullet (\mathbf{v}^* + \alpha\mathbf{g}) \leq b_{i,r}$.

$$\mathbf{a}_{i,r} \bullet \mathbf{g} = \mathbf{a}_{i,r} \bullet \Big(-(\mathbf{c} \bullet \mathbf{e})\mathbf{e} + \eta\mathbf{d} \Big) \leq 0$$

$$\mathbf{a}_{i,r} \bullet \mathbf{g} = -(\mathbf{c} \bullet \mathbf{e})(\mathbf{e} \bullet \mathbf{a}_{i,r}) + \eta(\mathbf{d} \bullet \mathbf{a}_{i,r}) \leq 0 \tag{B.28}$$

If $\mathbf{e} \bullet \mathbf{a}_{i,r} = 0$ then it is required that $\mathbf{d} \bullet \mathbf{a}_{i,r} \leq 0$, which is true by choice in Section B.5. When $\mathbf{e} \bullet \mathbf{a}_{i,r} < 0$, a limit exists if $\mathbf{d} \bullet \mathbf{a}_{i,r} > 0$:

$$\eta \leq \frac{(\mathbf{c} \bullet \mathbf{e})(\mathbf{e} \bullet \mathbf{a}_{i,r})}{\mathbf{d} \bullet \mathbf{a}_{i,r}} \tag{B.29}$$

if $\mathbf{d} \bullet \mathbf{a}_{i,r} \leq 0$ then any value of $\eta > 0$ will satisfy $\mathbf{a}_{i,r} \bullet \mathbf{g} \leq 0$.

The next set of limits on η is from external constraints where $h_j(\mathbf{v}^*) = 0$, the requirement is $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} < 0$ so that $\mathbf{g}$ points inwards. Therefore repeat Eq. B.28 but with $\nabla h_j(\mathbf{v}^*)$:

$$\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} = -(\mathbf{c} \bullet \mathbf{e})(\mathbf{e} \bullet \nabla h_j(\mathbf{v}^*)) + \eta(\mathbf{d} \bullet \nabla h_j(\mathbf{v}^*)) < 0 \tag{B.30}$$

If $j \in J_0$, i.e. $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$, then it is required that $\mathbf{d} \bullet \nabla h_j(\mathbf{v}^*) < 0$ so that Eq. B.30 is still held. This is true given the definition of $\mathbf{d}$ in Section B.5. For $j \in J_1$ a limit on η exists if $\mathbf{d} \bullet \nabla h_j(\mathbf{v}^*) > 0$:

$$\eta < \frac{(\mathbf{c} \bullet \mathbf{e})(\mathbf{e} \bullet \nabla h_j(\mathbf{v}^*))}{\mathbf{d} \bullet \nabla h_j(\mathbf{v}^*)} \tag{B.31}$$

if $\mathbf{d} \bullet \nabla h_j(\mathbf{v}^*) \leq 0$ then any value of $\eta > 0$ will satisfy $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} < 0$

Define η_{max} as the minimum value on the right hand side of the limits on η in Eqs. B.27, B.29, and B.31. Choose η in the interval $(0, \eta_{max})$ to define $\mathbf{g}$, e.g. $\eta = \eta_{max}/2$ is appropriate. If no limits for η exist then choose $\eta = 1$. The last set of steps is to find a limit on α greater than zero, thereby proving the existence of an element to C_r, and the termination of the algorithm. This process is a repetition of the analysis done for the simple case found in Eqs. B.21 and B.23, but here more detail is provided. It is already known for any value of α that $\mathbf{v}^* + \alpha\mathbf{g}$ is an element of $C(\mathbf{v}^*)$, and for all internal boundary constraints where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* = b_{i,r}$, $\mathbf{v}^* + \alpha\mathbf{g}$ is within that boundary. The first constraint limiting α are internal boundaries where $\mathbf{a}_{i,r} \bullet \mathbf{v}^* < b_{i,r}$:

$$\mathbf{a}_{i,r} \bullet (\mathbf{v}^* + \alpha\mathbf{g}) \leq b_{i,r}$$

If $\mathbf{a}_{i,r} \bullet \mathbf{g}$ is positive then there is a limit:

$$\alpha \leq \frac{b_{i,r} - \mathbf{a}_{i,r} \bullet \mathbf{v}^*}{\mathbf{a}_{i,r} \bullet \mathbf{g}} \tag{B.32}$$

if $\mathbf{a}_{i,r} \bullet \mathbf{g} \leq 0$ then any value of $\alpha > 0$ will suffice.

For external boundaries it is known that they are at most quadratic functions and can be expressed by the first three terms of its Taylor Series expansion:

$$h_j(\alpha) = h_j(\mathbf{v}^*) + \Big(\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}\Big)\alpha + \Big(\mathbf{g}^T H_j \mathbf{g}\Big)\alpha^2 \leq 0 \tag{B.33}$$

where H_j is the Hessian matrix, which is independent of $\mathbf{v}^*$. It is known that $h_j(\mathbf{v}) \leq 0$ forms a convex space, which requires $\mathbf{g}^T H_j \mathbf{g} \geq 0$ and H_j to be positive-semidefinite. This property will not be demonstrated here, but it is not too difficult to show from Eqs. B.2 and B.4. Therefore the limit on α depends on the following situations:

1. $h_j(\mathbf{v}^*) = 0$ and $\mathbf{g}^T H_j \mathbf{g} \geq 0$.

   This requires $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} < 0$, so that $h_j(\mathbf{v}^*) < 0$ for small α, and has been demonstrated to be true by the choice of η to satisfy Eq. B.31. A limit for α exists if $\mathbf{g}^T H_j \mathbf{g} > 0$:

   $$\alpha \leq -\frac{\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}}{\mathbf{g}^T H_j \mathbf{g}} \tag{B.34}$$

2. $h_j(\mathbf{v}^*) < 0$, $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} > 0$, and $\mathbf{g}^T H_j \mathbf{g} = 0$

   $$\alpha \leq -\frac{h_j(\mathbf{v}^*)}{\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}} \tag{B.35}$$

   A limit does not exist if $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} \leq 0$, as $h_j(\alpha)$ remains less than zero for all α.

3. $h_j(\mathbf{v}^*) < 0$ and $\mathbf{g}^T H_j \mathbf{g} > 0$

   $$\alpha \leq \frac{-\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g} + \sqrt{\Big(\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}\Big)^2 - 4h_j(\mathbf{v}^*)\Big(\mathbf{g}^T H_j \mathbf{g}\Big)}}{2\mathbf{g}^T H_j \mathbf{g}} \tag{B.36}$$

   This limit has to be positive due to the convexity of the space $h_j(\mathbf{v}) \leq 0$. Since $h_j(\mathbf{v}^*) < 0$ any direction from $\mathbf{v}^*$ that keeps $\mathbf{g}^T H_j \mathbf{g} > 0$ will eventually arrive at $h_j(\alpha) = 0$. In the above equation, the positive nature of the limit is guaranteed by:

   $$-4h_j(\mathbf{v}^*)\Big(\mathbf{g}^T H_j \mathbf{g}\Big) > 0$$

   $$\sqrt{\Big(\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}\Big)^2 - 4h_j(\mathbf{v}^*)\Big(\mathbf{g}^T H_j \mathbf{g}\Big)} > |-\nabla h_j(\mathbf{v}^*) \bullet \mathbf{g}|$$

Consequently a value is chosen for α satisfying Eqs. B.32, and B.34 to B.36, and is an element of C_r, proving that the local minimum of the new region is less than $\mathbf{c} \bullet \mathbf{v}^*$, and the algorithm will stop.

B.6.1 The Existence of d

In Sections B.5 and B.6, the existence of $\mathbf{d}$ in the direction of $\mathbf{e}$ is crucial to proving the algorithm terminates. The condition on $\mathbf{d}$ is that $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{d} < 0$ for all external constraints where $\nabla h_j(\mathbf{v}^*) \bullet \mathbf{e} = 0$ and $h_j(\mathbf{v}^*) = 0$. For a convex space, the only way for $\mathbf{d}$ not to exist is if there are two critical external boundaries $\nabla h_j(\mathbf{v}^*) = -\epsilon \nabla h_k(\mathbf{v}^*)$ for $\epsilon > 0$. This would also imply that there are two external boundaries in the vicinity of $\mathbf{v}^*$:

$$h_j(\mathbf{v}) = \mathbf{h} \bullet (\mathbf{v} - \mathbf{v}^*) \leq 0$$

$$h_k(\mathbf{v}) = -\epsilon \mathbf{h} \bullet (\mathbf{v} - \mathbf{v}^*) \leq 0$$

where $\mathbf{h} = \nabla h_j$. If such situation were to occur then the dimension of S_F, the complete feasible space, would be less than the number of variables. To avoid this situation, remove a variable from the problem by substituting $\mathbf{h} \bullet (\mathbf{v} - \mathbf{v}^*) = 0$ in all internal and external boundaries. Also remove the two external boundaries which caused this issue to arise.

B.7 NOTE ON TRANSITION RULES' PERFORMANCE

To the observant reader, it may seem that the rules for second choice of region, Section B.5, would always provide the optimal region to pick next. Then why is the first set of rules, Section B.3, not omitted altogether to make the algorithm more simple? It is true and a fair comment, but the reason is because the second choice rules are computational more onerous. The first set of rules just makes a judgement on the next region by which local boundaries it is on; there maybe quite a few rules, but only a small proportion is executed at each step. The second set of rules needs to check if the local minimum is global after most steps, requiring a set of linear equations to be solved, and then to perform an optimisation if the local minimum lies on multiple external boundaries to find $\mathbf{e}$. This will require significantly more computation time. Therefore it is my opinion that including the first set of rules would be the best approach. Knowing that from the four main examples of Chapter 5, never once were second choice rules ever required, and checks on global minimum, Section B.4, were only needed for low inertia scenarios of Example Four, it is a fair comment that the first choice rules will be faster.

It may be argued that if only using the second set of rules requires one less region to be checked, then the extra cost of adding each step may be justified. This may be true: consider how regions track by the two different sets of rules in Figure B.3. This uses Example 1 of Section 5.4, with the addition that the algorithm is applied for different starting regions. To solve this problem, only one starting region is required,

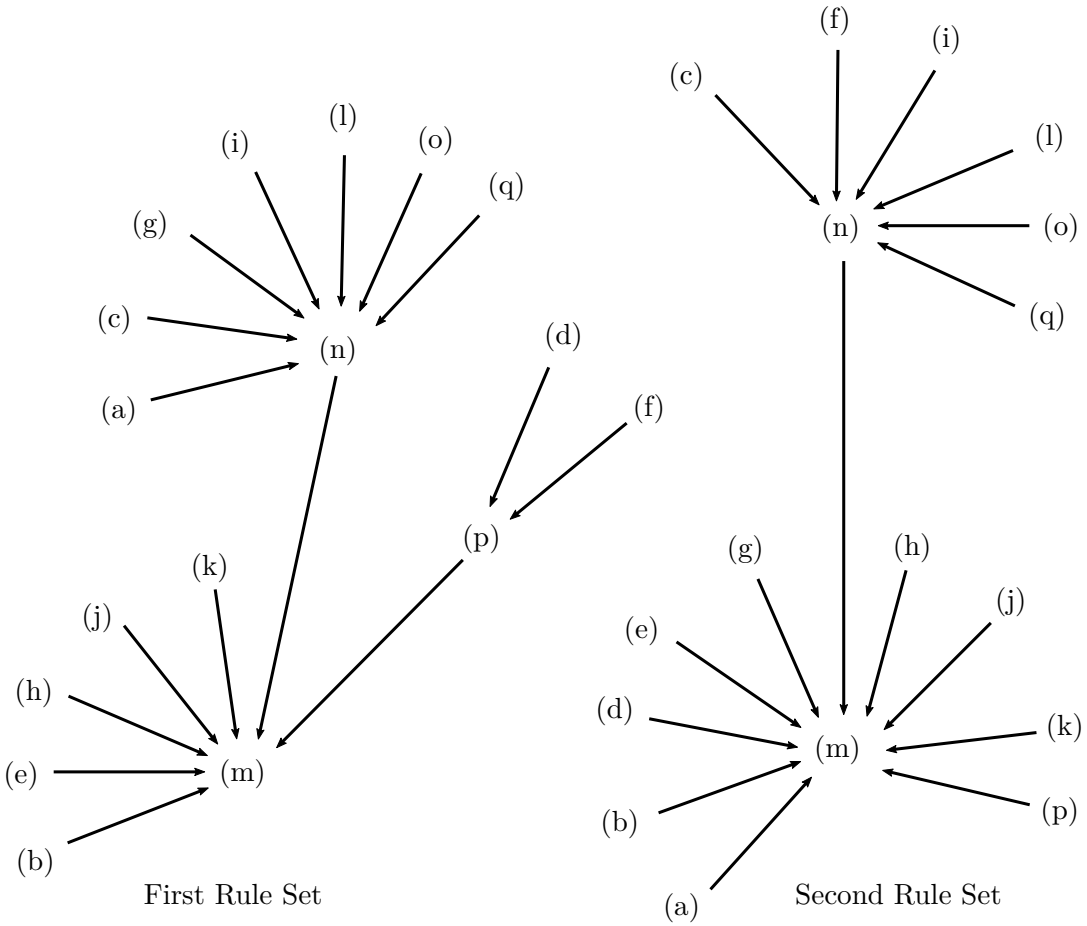


Figure B.3 The progression of the region tracking algorithm from region to region. Results show the tree structure after applying the first and second set of instructions to the first example, Section 5.4.1.

but to highlight the difference in performance it is necessary to show more. There is no difference in number of regions optimised when starting from region (c), as shown in Figure B.3. However if (a), (d), or (g) where chosen as the starting region then one less region is required with the second set of rules. Then for larger problems it is definitely possible that only using the second set of rules will be faster. To prove either way would need comparisons with larger problems including other constraints and offers.

It is not recommended that this testing be done, as it is anticipated that even faster algorithms will be found. The two approaches above suffer from having to find the local minimum of each region it is in, and then isolating its boundaries. Quicker methods would avoid this and only make shorter Newton Steps when approaching the global minimum, as shown in Figure B.4 where there are fewer Newton Steps in the faster approach. Allowing for Newton steps to cross boundaries will give it a significant advantage. Such a method is expected to be possible, but requires better familiarity with interior point methods, dual spaces, and how this applies with discontinuities. Therefore the more intuitively possible algorithm was developed for this research.

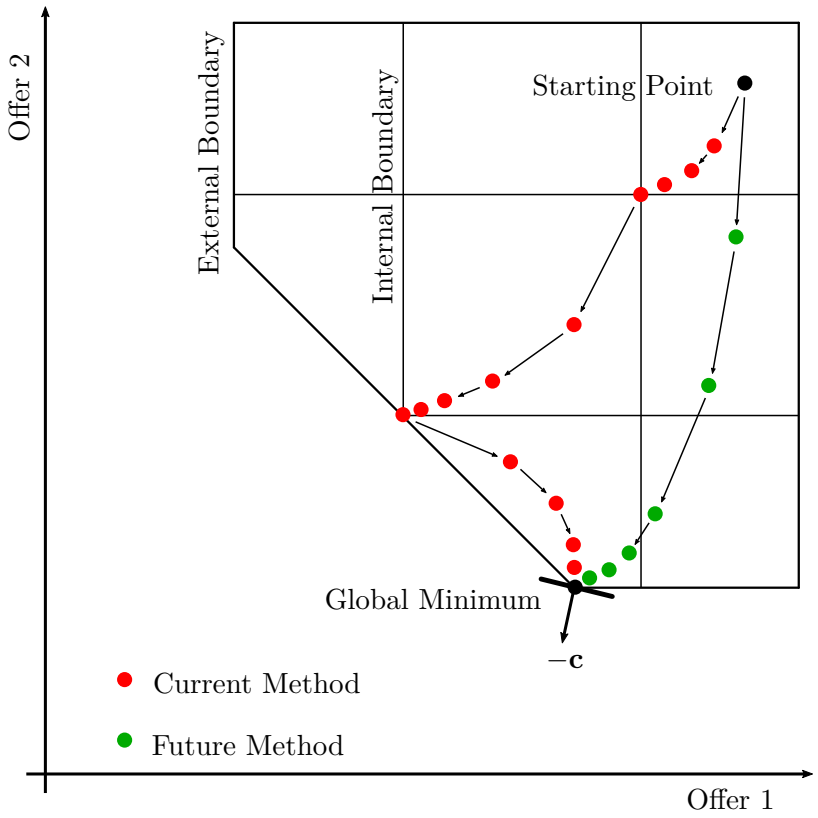


Figure B.4 A comparison between the current method of tracking from region to region and a future method that crosses internal boundaries in its Newton steps. The comparison is generalised for a two offer problem. Fewer green dots for the future method imply a faster solve time.

Appendix C

MARGINAL VALUE CALCULATION

In Section 5.4, results are shown for the optimisation process, a part of those solutions is commonly referred to as the shadow variables, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ of Eqs. 5.165 to 5.168. Usually some of these variables are important in setting prices in a market, such as the reserve price. However under the new formulation, these shadow variables will not be adequate for determining reserve prices. Therefore a general methodology is required to understand how the solutions change under variations in parameters.

Firstly some limitations have to be discussed:

- This analysis assumes that the global minimum has been found and it is known which region it is in, therefore in this Appendix the results, $\mathbf{y}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ only refer to that region, not to any of the regions solved before it.
- This analysis assumes that the solution is unique in all $\mathbf{y}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$, even though non-unique solutions are still valid results. It is anticipated that similar outcomes can be obtained in non-unique situations, but the concept of marginal will not apply to every variable.
- This analysis assumes that the global minimum does not lie on the boundary between regions, although in general it can. Similar results can be found but will require changes to the governing equations. It should be expected that in some instances and variables the marginal will be undefined when on the border.
- A general closed form equation for most marginal values is assumed not to exist, instead a methodology is given to calculating them.

A marginal is here defined as derivative of some part of the optimal solution, whether that be any of variables, e.g. $\mathbf{y}$, or any combination of variables, such as total cost, with any parameter of the problem. These parameters could be risk, R, when it is not a variable, or $t_{i,u}$, or any other parameter in the formulation. It could even be a parameter not explicitly defined in the formulation, but can be inserted into the equations with out changing the numeric form, like adding a parameter Δt to all time parameters and setting it to zero.

The optimal solution is the result from solving Eqs. 5.165 to 5.167. These equations can be expressed as a function, $\mathbf{f} = 0$:

$$\mathbf{f}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{bmatrix} \mathbf{c} + \left(G^T + 2H^T \otimes [\mathbf{y}]_{Ne}\right)\boldsymbol{\lambda} - \boldsymbol{\mu} \\ G\mathbf{y} + H\mathbf{y}^2 + \mathbf{m} \\ \mathbf{y} \otimes \boldsymbol{\mu} \end{bmatrix} \tag{C.1}$$

to reiterate, the meaning of the operator $\otimes$ is the element-wise multiplication; $\mathbf{y}^2$ is the element-wise squaring; and the operation $[\mathbf{y}]_{Ne}$ is the expansion of a vector between the brackets into a matrix by repeating it Ne times. The parameter of interest, generally defined as x, then can become a variable of the function, $\mathbf{f}(\mathbf{y}(x), \boldsymbol{\lambda}(x), \boldsymbol{\mu}(x), x) = 0$. Since it is known that function always has to equal zero to retain the optimal solution under any change in parameter x, then the derivative with respect to x has to equal zero as well:

$$\frac{d\mathbf{f}}{dx} = 0 \tag{C.2}$$

Applying the chain rule to evaluate the derivative, a series of equations is found that can be used to find the derivatives of $\mathbf{y}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$.

$$(2H^T\boldsymbol{\lambda}) \otimes \frac{d\mathbf{y}}{dx} + K^T\frac{d\boldsymbol{\lambda}}{dx} - \frac{d\boldsymbol{\mu}}{dx} = -\frac{\partial \mathbf{c}}{\partial x} - \frac{\partial K^T}{\partial x}\boldsymbol{\lambda} \tag{C.3}$$

$$K\frac{d\mathbf{y}}{dx} = -\frac{\partial G}{\partial x}\mathbf{y} - \frac{\partial H}{\partial x}\mathbf{y}^2 - \frac{\partial \mathbf{m}}{\partial x} \tag{C.4}$$

$$\boldsymbol{\mu} \otimes \frac{d\mathbf{y}}{dx} + \mathbf{y} \otimes \frac{d\boldsymbol{\mu}}{dx} = 0 \tag{C.5}$$

where $K^T = G^T + 2H^T \otimes [\mathbf{y}]_{Ne}$. The number of equations can be reduced by substituting Eq. C.5 into Eq. C.3 after multiplying each i^{th} equation of Eq. C.3 by y_i:

$$\left(\boldsymbol{\mu} + \mathbf{y} \otimes (2H^T\boldsymbol{\lambda})\right) \otimes \frac{d\mathbf{y}}{dx} + ([\mathbf{y}]_{Ne} \otimes K^T)\frac{d\boldsymbol{\lambda}}{dx} = -\mathbf{y} \otimes \left(\frac{\partial \mathbf{c}}{\partial x} + \frac{\partial K^T}{\partial x}\boldsymbol{\lambda}\right) \tag{C.6}$$

Therefore the derivatives are found by solving the following matrix equation:

$$\left[\begin{array}{c|c} \operatorname{diag}(\boldsymbol{\mu} + \mathbf{y} \otimes (2H^T\boldsymbol{\lambda})) & [\mathbf{y}]_{Ne} \otimes K^T \\ \hline K & \mathbf{0}_{Ne \times Ne} \end{array}\right] \left[\begin{array}{c} \frac{d\mathbf{y}}{dx} \\ \hline \frac{d\boldsymbol{\lambda}}{dx} \end{array}\right] = -\left[\begin{array}{c} \mathbf{y} \otimes \left(\frac{\partial \mathbf{c}}{\partial x} + \frac{\partial K^T}{\partial x}\boldsymbol{\lambda}\right) \\ \hline \frac{\partial G}{\partial x}\mathbf{y} + \frac{\partial H}{\partial x}\mathbf{y}^2 + \frac{\partial \mathbf{m}}{\partial x} \end{array}\right] \tag{C.7}$$

The derivative for $\boldsymbol{\mu}$ can then be found through Eq. C.3. To simplify describing this matrix equation, Eq. C.7 is simplified to $\Psi\boldsymbol{\omega} = -\boldsymbol{\gamma}$. Next two important marginals

are expressed: the total reserve dispatched, and the total cost. The first of these is:

$$\mathbf{1}^T \frac{d\mathbf{p}}{dx} + \mathbf{1}^T \frac{d\mathbf{u}}{dx} \tag{C.8}$$

where $\mathbf{p} = [p_1, \ldots, p_{Np}]^T$ and $\mathbf{u} = [u_1, \ldots, u_{Nu}]^T$. The marginal change in cost to changes in parameter x is:

$$\frac{\partial \mathbf{c}^T}{\partial x}\mathbf{y} + \mathbf{c}^T \frac{d\mathbf{y}}{dx} \tag{C.9}$$

Finding a specific result, e.g. a marginal for changing risk, R, is matter of finding the partial derivatives of $\boldsymbol{\gamma}$. For the example of risk, it is only $\frac{\partial \mathbf{m}}{\partial R}$ that is not zero. It is noticed that the matrix Ψ is independent of the parameter of interest, and is characteristic to the problem. If the marginals of multiple parameters are required then it may be computationally efficient to find the inverse of Ψ.

REFERENCES

Abeyratne, S. (2007), 'Effect of wind generation on management of frequency excursions: Investigation 5 of WGIP', Tech. rep., Transpower.

Abhyankar, S., Geng, G., Anitescu, M., Wang, X. and Dinavahi, V. (2017), 'Solution techniques for transient stability-constrained optimal power flow: Part i', *IET Generation, Transmission Distribution*, Vol. 11, No. 12, pp. 3177–3185.

Ackermann, T., Prevost, T., Vittal, V., Roscoe, A.J., Matevosyan, J. and Miller, N. (2017), 'Paving the way: A future without inertia is closer than you think', *IEEE Power and Energy Magazine*, Vol. 15, No. 6, pp. 61–69.

Ackermann, T. (2012), *Wind power in power systems*, Wiley, Chichester, West Sussex; Hoboken, N.J;, 2nd ed.

AEMC (2017), *The Frequency Operating Standards*, `https://www.aemc.gov.au/regulation/electricity-guidelines-and-standards`, Australian Energy Market Commission, Australia.

AEMO (2015), *Guide to Ancillary Services in the National Electricity Market*, `https://www.aemo.com.au/-/media/Files/PDF/Guide-to-Ancillary-Services-in-the-National-Electricity-Market.pdf`, Australian Energy Market Operator.

AEMO (2017a), *Black System South Australia 28 September 2016*, `https://www.aemo.com.au/-/media/Files/Electricity/NEM/Market_Notices_and_Events/Power_System_Incident_Reports/2017/Integrated-Final-Report-SA-Black-System-28-September-2016.pdf`, Australian Energy Market Operator.

AEMO (2017b), *Causer Pays Procedure: Determination of Contribution Factors for Regulation FCAS Cost Recover*, `https://www.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/Ancillary_Services/20170106-Causer-Pays-Procedure.pdf`, Australian Energy Market Operator.

AEMO (2017c), *Fast Frequency Response in the NEM*, `https://www.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/Reports/2017/FFR-Working-Paper---Final.pdf`, Australian Energy Market Operator.

AEMO (2018a), 'Hornsdale Wind Farm 2 FCAS trial', Tech. rep., Australian Electricity Market Regulator.

AEMO (2018b), *Inertia Requirements Methodology: Inertia Requirements and Shortfalls*, http://energylive.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/System-Security-Market-Frameworks-Review/2018/Inertia_Requirements_Methodology_PUBLISHED.pdf, Australian Energy Market Operator.

AEMO (2018c), *Power System Frequency Risk Preview Report*, https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/PSFRR/2018_Power_System_Frequency_Risk_Review-Final_Report.pdf, Australian Energy Market Operator.

Ahmadi, H. and Ghasemi, H. (2014), 'Security-constrained unit commitment with linearized system frequency limit constraints', *IEEE Transactions on Power Systems*, Vol. 29, No. 4, pp. 1536–1545.

Akhmatov, V. (2003), *Analysis of dynamic behaviour of electric power systems with large amount of wind power*, PhD thesis, Technical University of Denmark.

Anderson, P.M. and Fouad, A.A. (2003), *Power system control and stability*, IEEE Press, Piscataway, N.J, 2nd ed.

Aneke, M. and Wang, M. (2016), 'Energy storage technologies and real life applications: A state of the art review', *Applied Energy*, Vol. 179, pp. 350 – 377.

Appleby, S. and Rositano, P. (2019), 'Addressing the System Strength Gap in SA', Tech. rep., ElectraNet, Australia.

Ashton, P.M., Saunders, C.S., Taylor, G.A., Carter, A.M. and Bradley, M.E. (2015), 'Inertia estimation of the GB power system using synchrophasor measurements', *IEEE Transactions on Power Systems*, Vol. 30, No. 2, pp. 701–709.

Asmine, M., Langlois, C. and Aubut, N. (2018), 'Inertial response from wind power plants during a frequency disturbance on the hydro-quebec system: event analysis and validation', *IET Renewable Power Generation*, Vol. 12, No. 5, pp. 515–522.

Barbour, E., Wilson, I.G., Radcliffe, J., Ding, Y. and Li, Y. (2016), 'A review of pumped hydro energy storage development in significant international electricity markets', *Renewable and Sustainable Energy Reviews*, Vol. 61, pp. 421 – 432.

Bird, L., Lew, D., Milligan, M., Carlini, E.M., Estanqueiro, A., Flynn, D., Gomez-Lazaro, E., Holttinen, H., Menemenlis, N., Orths, A., Eriksen, P.B., Smith, J.C., Soder, L., Sorensen, P., Altiparmakis, A., Yasuda, Y. and Miller, J. (2016), 'Wind and solar energy curtailment: A review of international experience', *Renewable and Sustainable Energy Reviews*, Vol. 65, pp. 577 – 586.

Brown, B.J. (1998), *Power System Low Frequency Limitations*, https://web.archive.org/web/20040611090955/http://www.gsp.co.nz:80/welcome.html, Security Develepment Working Group, New Zealand.

Brown, R.W. and Scott-Dye, H.N. (2012), 'Te uku wind farm - planning and operation of a deeply embedded power plant with advanced ancillary services', In *CIGRE Session 43*, Paris, France.

Bull, B. (2010), 'Correlation between wind generation output and hydro inflows', Tech. rep., Electricity Commission, New Zealand.

CAE (1993), *Reliability of Electricity Supply*, Centre for Advanced Engineering, University of Canterbury, New Zealand.

Chávez, H., Baldick, R. and Sharma, S. (2014), 'Governor rate-constrained OPF for primary frequency control adequacy', *IEEE Transactions on Power Systems*, Vol. 29, No. 3, pp. 1473–1480.

Chen, H., Baker, S., Benner, S., Berner, A. and Liu, J. (2017), 'PJM integrates energy storage: Their technologies and wholesale products', *IEEE Power and Energy Magazine*, Vol. 15, No. 5, pp. 59–67.

Chong, E.K.P. and Zak, S.H. (2013), *An Introduction to Optimization*, Wiley, Hoboken, New Jersey, 4th ed.

CIGRE (1999), 'Large frequency disturbances: analysis and modeling needs', In *IEEE Power Engineering Society. 1999 Winter Meeting (Cat. No.99CH36233)*, Vol. 1, pp. 554–558 vol.1.

Cramton, P. and Ockenfels, A. (2011), *Economics and design of capacity markets for the power sector.*

CRU (2018), 'Rate of Change of Frequency Project Quarterly Report for Q3 2018', `https://www.cru.ie/wp-content/uploads/2018/11/CRU18237-RoCoF-Project-Quarterly-Report-Q3-2018.pdf`, Commission for Regulation of Utilities, Ireland.

Dehghanpour, K. and Afsharnia, S. (2015), 'Electrical demand side contribution to frequency control in power systems: a review on technical aspects', *Renewable and Sustainable Energy Reviews*, Vol. 41, pp. 1267 – 1276.

Díaz-González, F., Hau, M., Sumper, A. and Gomis-Bellmunt, O. (2014), 'Participation of wind power plants in system frequency control: Review of grid code requirements and control methods', *Renewable and Sustainable Energy Reviews*, Vol. 34, pp. 551 – 564.

Doherty, R., Lalor, G. and O'Malley, M. (2005), 'Frequency control in competitive electricity market dispatch', *IEEE Transactions on Power Systems*, Vol. 20, No. 3, pp. 1588–1596.

EA (2016), *Ancillary services procurment plan*, `https://www.ea.govt.nz/code-and-compliance/the-code/amendments/amendents-to-documents-incorporated-by-reference/`, Electricity Authority, New Zealand.

EA (2017a), *Efficient procurement of extended reserve*, `https://www.ea.govt.nz/development/work-programme/risk-management/efficient-procurement-extended-reserve/development/decision-paper-on-efficient-procurement-of-extended-reserve/`, Electricity Authority, New Zealand.

EA (2017b), *Normal Frequency Management: Strategic Review Information Paper*, https://www.ea.govt.nz/development/work-programme/risk-management/normal-frequency-management-strategic-review/consultations/#c16388, Electricity Authority, New Zealand.

EA (2018a), 'Electricity in New Zealand', Tech. rep., Electricity Authority, https://www.ea.govt.nz/about-us/media-and-publications/electricity-nz/.

EA (2018b), *Electricity Industry Participation Code*, https://www.ea.govt.nz/code-and-compliance/the-code/, Electricity Authority, New Zealand.

EA (2018c), *Normal Frequency Management: Decision Paper*, https://www.ea.govt.nz/development/work-programme/risk-management/normal-frequency-management-strategic-review/development/decision-paper-on-normal-frequency-management/, Electricity Authority, New Zealand.

EA (2018d), *Review of instantaneous reserve markets project: Decision Paper*, https://www.ea.govt.nz/development/work-programme/risk-management/ufm-reserve-arrangements/development/review-of-instantaneous-reserve-markets-project/, Electricity Authority, New Zealand.

EirGrid (2015), 'Eirgrid Grid Code Version 6.0', http://www.eirgridgroup.com/site-files/library/EirGrid/GridCodeVersion6.pdf, Clause CC.7.3.1.1.d.

EirGrid, SEMO and SONI (2014), *DS3 System Services: Portfolio Capability Analysis*.

Ela, E., Gevorgian, V., Tuohy, A., Kirby, B., Milligan, M. and O'Malley, M. (2014), 'Market designs for the primary frequency response ancillary service part i: Motivation and design', *IEEE Transactions on Power Systems*, Vol. 29, No. 1, pp. 421–431.

Ellison, J.F., Tesfatsion, L.S., Loose, V.W. and Byrne, R.H. (2012), *Project Report: A Survey of Operating Reserve Markets in U.S. ISO/RTO-managed Electric Energy Regions*, Sandia National Laboratories, USA.

Elsgolc, L. (2012), *Calculus of Variations*, Dover Books on Mathematics, Dover Publications.

ENERNOC (2015), *Our electricity system is changing - what this means to the Smart Grid Forum?*, Working paper to the Smart Grid Forum, MBIE.

Engelken, S., Mendonca, A. and Fischer, M. (2017), 'Inertial response with improved variable recovery behaviour provided by type 4 WTs', *IET Renewable Power Generation*, Vol. 11, No. 3, pp. 195–201.

ENTSOE (2017), 'Future System Inertia 2', Tech. rep., European Network of Transmission System Operators, https://www.entsoe.eu/publications/system-operations-reports/#nordic, on behalf of Nordic Power System.

ERCOT (2018), *Inertia: Basic Concepts and Impacts on the ERCOT Grid*, http://www.ercot.com/content/wcm/lists/144927/Inertia_Basic_Concepts_Impacts_On_ERCOT_v0.pdf, Electric Reliability Council of Texas.

Evans, L. (2017), 'The electricity spot market: Is it future proof?', *The Electricity Journal*, Vol. 30, No. 2, pp. 25–29.

Evans, L.T. and Meade, R.B. (2005), *Alternating currents or counter-revolution?: contemporary electricity reform in New Zealand*, Victoria University Press, Wellington [N.Z.].

FERC (2018), *Electric Storage Participation in Markets Operated by Regional Transmission Organizations and Independent System Operators*, Order No. 841, Federal Energy Regulatory Commission.

Fischer, M., Engelken, S., Mihov, N. and Mendonca, A. (2016), 'Operational experiences with inertial response provided by type 4 wind turbines', *IET Renewable Power Generation*, Vol. 10, No. 1, pp. 17–24.

FSWG (2001), 'Cost Benefit Analysis Report', Tech. rep., Grid Security Committee, https://web.archive.org/web/20040611090955/http://www.gsp.co.nz:80/welcome.html, Frequency Standards Working Group, New Zealand.

Geng, G., Abhyankar, S., Wang, X. and Dinavahi, V. (2017), 'Solution techniques for transient stability-constrained optimal power flow: Part ii', *IET Generation, Transmission Distribution*, Vol. 11, No. 12, pp. 3186–3193.

Geth, F., Brijs, T., Kathan, J., Driesen, J. and Belmans, R. (2015), 'An overview of large-scale stationary electricity storage plants in europe: Current status and new developments', *Renewable and Sustainable Energy Reviews*, Vol. 52, pp. 1212 – 1227.

González, A., Goikolea, E., Barrena, J.A. and Mysyk, R. (2016), 'Review on supercapacitors: Technologies and materials', *Renewable and Sustainable Energy Reviews*, Vol. 58, pp. 1189 – 1206.

Greve, T., Teng, F., Pollitt, M.G. and Strbac, G. (2018), 'A system operator's utility function for the frequency response market', *Applied Energy*, Vol. 231, pp. 562 – 569.

Guggilam, S.S., Zhao, C., Dall'Anese, E., Chen, Y.C. and Dhople, S.V. (2018), 'Optimizing DER participation in inertial and primary-frequency response', *IEEE Transactions on Power Systems*, Vol. 33, No. 5, pp. 5194–5205.

Hartl, R.F., Sethi, S.P. and Vickson, R.G. (1995), 'A survey of the maximum principles for optimal control problems with state constraints', *SIAM Review*, Vol. 37, No. 2, pp. 181–218.

He, W. and Wang, J. (2018), 'Optimal selection of air expansion machine in compressed air energy storage: A review', *Renewable and Sustainable Energy Reviews*, Vol. 87, pp. 77 – 95.

Hu, X., Zou, C., Zhang, C. and Li, Y. (2017), 'Technological developments in batteries: A survey of principal roles, types, and management needs', *IEEE Power and Energy Magazine*, Vol. 15, No. 5, pp. 20–31.

IEEE PES (2011), 'IEEE guide for the application of turbine governing systems for hydroelectric generating units - redline', *IEEE Std 1207-2011 (Revision to IEEE Std 1207-2004) - Redline*, pp. 1–139.

Joos, M. and Staffell, I. (2018), 'Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany', *Renewable and Sustainable Energy Reviews*, Vol. 86, pp. 45 – 65.

Kazaglis, A., Ward, J., Evans, S., Sammon, P. and Kemp, L. (2017), 'Net zero in New Zealand: Scenarios to achieve domestic emissions neutrality in the second half of the century', Tech. rep., Vivid Economics.

Kirby, B. (2014), *Potential New Ancillary Services: Developments of interest to generators*, `http://www.consultkirby.com/files/01B2_Kirby_Paper_-_Potential_New_Ancillary_Services.pdf`.

Kong, Y., Kong, Z., Liu, Z., Wei, C., Zhang, J. and An, G. (2017), 'Pumped storage power stations in China: The past, the present, and the future', *Renewable and Sustainable Energy Reviews*, Vol. 71, pp. 720 – 731.

Koritarov, V., Guzowski, L., Feltes, J., Kazachkov, Y., Gong, B., Trouille, B. and Donalek, P. (2013a), 'Modeling adjustable speed pumped storage hydro units employing doubly-fed induction machines', Tech. rep., Argonne National Laboratory.

Koritarov, V., Guzowski, L., Feltes, J., Kazachkov, Y., Lam, B., Grande-Moran, C., Thomann, G., Eng, L., Trouille, B. and Donalek, P. (2013b), 'Review of existing hydroelectric turbine-governor simulation models', Tech. rep., Argonne National Laboratory.

Krause, P.C., Wasynczuk, O., Sudhoff, S.D. and Pekarek, S. (2013), *Analysis of electric machinery and drive systems*, Wiley, Hoboken, New Jersey, 3rd ed.

Kroposki, B., Johnson, B., Zhang, Y., Gevorgian, V., Denholm, P., Hodge, B. and Hannegan, B. (2017), 'Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy', *IEEE Power and Energy Magazine*, Vol. 15, No. 2, pp. 61–73.

Kundur, P., Balu, N.J. and Lauby, M.G. (1994), *Power system stability and control*, McGraw-Hill, New York.

Kundur, P., Paserba, J., Ajjarapu, V., Andersson, G., Bose, A., Canizares, C., Hatziargyriou, N., Hill, D., Stankovic, A., Taylor, C., Cutsem, T.V. and Vittal, V. (2004), 'Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions', *IEEE Transactions on Power Systems*, Vol. 19, No. 3, pp. 1387–1401.

Lee, Y. and Baldick, R. (2013), 'A frequency-constrained stochastic economic dispatch model', *IEEE Transactions on Power Systems*, Vol. 28, No. 3, pp. 2301–2312.

Lewis, F., Syrmos, V. and Syrmos, V. (1995), *Optimal Control*, A Wiley-interscience publication, Wiley.

Li, W., Du, P. and Lu, N. (2018), 'Design of a new primary frequency control market for hosting frequency response reserve offers from both generators and loads', *IEEE Transactions on Smart Grid*, Vol. 9, No. 5, pp. 4883–4892.

Liu, C. and Du, P. (2018), 'Participation of load resources in day-ahead market to provide primary-frequency response reserve', *IEEE Transactions on Power Systems*, Vol. 33, No. 5, pp. 5041–5051.

McQueen, D. (2016), *Quantifying the benefits from the spatial diversification of wind power in New Zealand*, PhD thesis, University of Canterbury.

Miller, N.W. and Pajic, S. (2016), 'Diverse fast frequency response services in systems with declining synchronous inertia', In *15th Wind Integration Workshop, Vienna, Austria*, .

Miller, R. (2014), 'Area Under the Curve (AUTC) - A new approach to Instantaneous Reserves modelling', Electricity Authority, New Zealand.

Mousavi G, S., Faraji, F., Majazi, A. and Al-Haddad, K. (2017), 'A comprehensive review of flywheel energy storage system technology', *Renewable and Sustainable Energy Reviews*, Vol. 67, pp. 477 – 490.

NERC (2018), *Reliability Standards for the Bulk Electric Systems of North America*, `https://www.nerc.com/pa/Stand/Reliability%20Standards%20Complete%20Set/RSCompleteSet.pdf`, North American Electric Reliability Corporation.

New Zealand Government (2019), *Climate Change Response (Zero Carbon) Amendment Bill*, currently passed the first reading.

NG (2016), *Enhanced Frequency Response Market Information Report*, `https://www.nationalgrideso.com/sites/eso/files/documents/EFR%20Market%20Information%20Report%20v1.pdf`, National Grid, United Kingdom.

NZX (2017), 'Extended Reserve Selection Methodology', Tech. rep., New Zealand Exchange.

O'Sullivan, J., Rogers, A., Flynn, D., Smith, P., Mullane, A. and O'Malley, M. (2014), 'Studying the maximum instantaneous non-synchronous generation in an island system: Frequency stability challenges in Ireland', *IEEE Transactions on Power Systems*, Vol. 29, No. 6, pp. 2943–2951.

Palermo, J. (2016), 'International review of frequency control adaption', Tech. rep., DGA Consulting, AEMO, `https://www.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/Reports/2016/FPSS---International-Review-of-Frequency-Control.pdf`.

Pelletier, M.A., Phethean, M.E. and Nutt, S. (2012), 'Grid code requirements for artificial inertia control systems in the New Zealand power system', In *2012 IEEE Power and Energy Society General Meeting*, pp. 1–7.

Peng, C., Zou, J. and Lian, L. (2017), 'Dispatching strategies of electric vehicles participating in frequency regulation on power grid: A review', *Renewable and Sustainable Energy Reviews*, Vol. 68, pp. 147 – 152.

Phethean, M., Bickers, G., Di Somma, D. and McKinnon, H. (2015a), 'New Instantaneous Reserve Products', Tech. rep., Transpower.

Phethean, M., Pat, D., Adam, N. and Bartley, J. (2015b), 'Frequency Keeping Control Trial Technical Review Report', Tech. rep., Transpower.

Philpott, A., Read, G., Batstone, S. and Miller, A. (2018), 'The New Zealand electricity market: challenges of a renewable energy system', Tech. rep., Electric Power Optimization Centre, `www.epoc.org.nz/publications.html`, New Zealand.

Poletti, S. (2018), 'Market power in the NZ wholesale market 2010 to 2016', Tech. rep., Energy Centre, University of Auckland, `https://www.business.auckland.ac.nz/en/about/our-research/bs-research-institutes-and-centres/energy-centre/publications-22.html`.

Pourbeik, P., Kundur, P.S. and Taylor, C.W. (2006), 'The anatomy of a power grid blackout - root causes and dynamics of recent major blackouts', *IEEE Power and Energy Magazine*, Vol. 4, No. 5, pp. 22–29.

Rebours, Y. (2008), *A comprehensive assessment of markets for frequency and voltage control ancillary services*, PhD thesis, University of Manchester.

Rehman, S., Al-Hadhrami, L.M. and Alam, M.M. (2015), 'Pumped hydro energy storage system: A technological review', *Renewable and Sustainable Energy Reviews*, Vol. 44, pp. 586 – 598.

Restrepo, J.F. and Galiana, F.D. (2005), 'Unit commitment with primary frequency regulation constraints', *IEEE Transactions on Power Systems*, Vol. 20, No. 4, pp. 1836–1842.

Roberts, C. (2018), 'Review of international grid codes', Tech. rep., Lawrence Berkeley National Laboratory.

Ross, A. (2001), *Performance of Thermal Plant at Low Frequencies*, `https://web.archive.org/web/20040611090955/http://www.gsp.co.nz:80/welcome.html`, PB Power, New Zealand.

Schipper, J., Wood, A., Edwards, C. and Miller, A. (2019), 'Recommendations for Ancillary Service Markets under High Penetrations of Wind Generation in New Zealand', Tech. rep., EPECentre.

Schweppe, F.C. (1988), *Spot pricing of electricity*, Kluwer Academic, Boston.

SEM (2013), 'DS3 System Services Technical Definitions - Decision Paper', Tech. rep., Single Electricity Market, tech. Rep. SEM-13-098.

Shair, J., Xie, X., Wang, L., Liu, H., Liu, W. and He, J. (2019), 'Overview of emerging subsynchronous oscillations in practical wind power systems', *Renewable and Sustainable Energy Reviews*, Vol. 99, pp. 159–168.

Singarao, V.Y. and Rao, V.S. (2016), 'Frequency responsive services by wind generation resources in United States', *Renewable and Sustainable Energy Reviews*, Vol. 55, pp. 1097 – 1108.

Singh, R.R., Chelliah, T.R. and Agarwal, P. (2014), 'Power electronics in hydro electric energy systems: A review', *Renewable and Sustainable Energy Reviews*, Vol. 32, pp. 944 – 959.

Smith, P.K. (1989), 'New load shedding technique avoids power disruption', *Marketing Matters, Electricorp Marketing NZ*.

SO (2007), 'Wind Generation Investigation Project, Investigations 1 to 9', Tech. rep., System Operator, Transpower, https://www.ea.govt.nz/about-us/what-we-do/our-history/archive/dev-archive/consultations/power-systems-and-common-quality-consultations/2005/wind-generation-investigation-project/.

SO (2014), 'TASC 033 Report', Tech. rep., Transpower, https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/TASC%20033%20report.pdf.

SO (2016), *GL-EA-010 Companion Guide for Testing of Assets*, https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/GL-EA-010_Companion%20Guide%20for%20Testing%20of%20Assets.pdf, System Operator, Transpower, New Zealand.

Sokoler, L.E., Vinter, P., Baerentsen, R., Edlund, K. and Jorgensen, J.B. (2016), 'Contingency-constrained unit commitment in meshed isolated power systems', *IEEE Transactions on Power Systems*, Vol. 31, No. 5, pp. 3516–3526.

Strbac, G., Pudjianto, D., Shakoor, A.A., Castro, M.J., Waipara, G. and Telfar, G. (2008), 'The system impacts and costs of integrating wind power in New Zealand', Tech. rep., Imperial College London and Meridian Energy.

Teeuwsen, S.P., Love, G. and Sherry, R. (2013), '1400 MW New Zealand HVDC upgrade: Introducing power modulation controls and round power mode', In *IEEE Power and Energy Society General Meeting*, .

Teng, F., Trovato, V. and Strbac, G. (2016), 'Stochastic scheduling with inertia-dependent fast frequency response requirements', *IEEE Transactions on Power Systems*, Vol. 31, No. 2, pp. 1557–1566.

Transpower (2014), *Credible Event Review*, https://www.transpower.co.nz/system-operator/key-documents/reports#2014, New Zealand.

Transpower (2017a), *Battery Storage in New Zealand: Discussion Document*, https://www.transpower.co.nz/sites/default/files/publications/resources/Battery%20Storage%20in%20New%20Zealand.pdf.

Transpower (2017b), *Extended Reserve Technical Requirements Schedule*, https://www.transpower.co.nz/system-operator/electricity-market/extended-reserve.

Transpower (2017c), *Policy Statement*, https://www.ea.govt.nz/code-and-compliance/the-code/documents-incorporated-into-the-code-by-reference/, New Zealand.

Transpower (2017d), 'Solar PV in New Zealand, Effect of Solar PV on Frequency Management in New Zealand, Effect of Solar PV on Generation Dispatch in New Zealand, Effect of Solar PV on Transient Stability in New Zealand Power System', Tech. rep., Transpower.

Transpower (2018a), *GL-OC-209 SPD Schedule Inputs Version 5.0*, https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/GL-OC-209%20SPD%20Schedule%20Inputs.pdf.

Transpower (2018b), *RMT Functional Specification Version 6.0*, https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/RMT_Specification_V6_issue.pdf, New Zealand.

Transpower (2018c), *SPD Model Formulation Version 11.2*, https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/SPD_Model_Formulation_v11.2.pdf, New Zealand.

Wang, L., Xie, X., Dong, X., Liu, Y. and Shen, H. (2018), 'Real-time optimisation of short-term frequency stability controls for a power system with renewables and multi-infeed HVDCs', *IET Renewable Power Generation*, Vol. 12, No. 13, pp. 1462–1469.

Wen, Y., Li, W., Huang, G. and Liu, X. (2016), 'Frequency dynamics constrained unit commitment with battery energy storage', *IEEE Transactions on Power Systems*, Vol. 31, No. 6, pp. 5115–5125.

Wolak, F.A. (2009), 'An Assessment of the Peformence of the New Zealand Wholesale Electricity Market', Tech. rep., Commerce Commission, New Zealand.

Wright, S.J. (1997), *Primal-dual interior-point methods*, Society for Industrial and Applied Mathematics, Philadelphia.

Yaramasu, V., Wu, B., Sen, P.C., Kouro, S. and Narimani, M. (2015), 'High-power wind energy conversion systems: State-of-the-art and emerging technologies', *Proceedings of the IEEE*, Vol. 103, No. 5, pp. 740–788.

Young, L.M. (2009), 'A collation of international policies for under-frequency load shedding', Tech. rep., Transpower.

Zhang, Y., Gevorgian, V., Wang, C., Lei, X., Chou, E., Yang, R., Li, Q. and Jiang, L. (2017), 'Grid-level application of electrical energy storage: Example use cases in the United States and China', *IEEE Power and Energy Magazine*, Vol. 15, No. 5, pp. 51–58.